

# An Integer Programming Based Bound for Locally Repairable Codes

Anyu Wang and Zhifang Zhang

Key Laboratory of Mathematics Mechanization, NCMIS

Academy of Mathematics and Systems Science, CAS, Beijing, China

Email: {wanganyu, zfz}@amss.ac.cn

## Abstract

The locally repairable code (LRC) studied in this paper is an  $[n, k]$  linear code of which the value at each coordinate can be recovered by a linear combination of at most  $r$  other coordinates. The central problem in this work is to determine the largest possible minimum distance for LRCs. First, an integer programming based upper bound is derived for any LRC. Then by solving the programming problem under certain conditions, an explicit upper bound is obtained for LRCs with parameters  $n_1 > n_2$ , where  $n_1 = \left\lceil \frac{n}{r+1} \right\rceil$  and  $n_2 = n_1(r+1) - n$ . Finally, an explicit construction for LRCs attaining this upper bound is presented over the finite field  $\mathbb{F}_{2^m}$ , where  $m \geq n_1 r$ . Based on these results, the largest possible minimum distance for all LRCs with  $r \leq \sqrt{n} - 1$  has been definitely determined, which is of great significance in practical use.

## I. INTRODUCTION

In distributed storage systems, redundancy must be introduced to protect data against device failures. The simplest form of redundancy is *replication*. But it is extremely inefficient due to its large storage overhead, namely,  $c$  copies of the data have to be stored to guarantee  $(c-1)$ -erasure tolerance. To improve the storage efficiency, *erasure codes* are employed in distributed storage systems, such as Windows Azure [5], Facebook's Hadoop cluster [14], etc, where the original data are divided into  $k$  equal-size fragments and then encoded into  $n$  fragments ( $n > k$ ) stored in  $n$  different nodes. The fault tolerance property of the erasure code ensures that the system can tolerate up to  $d-1$  node failures, where  $d$  is the minimum distance of the erasure code. Particularly, the MDS code is a kind of erasure code that attains the optimal minimum distance with respect to the Singleton bound and thus provides the highest level of fault tolerance for given storage overhead. But the MDS code is still inefficient for distributed storage systems because of the disk I/O complexity it causes in the *node repair* issue. Specifically, when an  $[n, k]$  MDS code is employed, repairing a failed node usually needs the access of  $k$  other survival nodes, which entails too much complexity in contrast with the amount of data to be repaired.

To improve this, Gopalan et al. [3], Oggier et al. [7], and Papailiopoulos et al. [10] introduced *repair locality* for erasure codes. The  $i$ th coordinate of a code has repair locality  $r$  if the value at this coordinate can be recovered by accessing at most  $r$  other coordinates. In more detail, a code is said to have *information locality* if the locality  $r$  is ensured for each coordinate in an information set containing information symbols, e.g., systematic coordinates in a linear systematic code. Alternatively, a code is said to have *all symbol locality* if the locality  $r$  is ensured for all coordinates. In this paper we call an  $[n, k]$  linear code with all symbol locality  $r$  as a *locally repairable code* (LRC). When  $r \ll k$  it greatly reduces the disk I/O complexity for repair.

Considering the fault tolerance level, the minimum distance is also an important metric for LRCs. Gopalan et al. [3] first derived the following upper bound for codes with information locality:

$$d \leq n - k + 1 - \left( \left\lceil \frac{k}{r} \right\rceil - 1 \right) \quad (1)$$

which is a tight bound by the construction of pyramid codes [4]. Although the bound (1) certainly holds for LRCs, it is not tight in many cases. The results in [3] pointed out that when  $(r+1) \nmid n$  and  $r \mid k$  the

bound (1) cannot be attained for codes with all symbol locality, and for those attaining this bound only the existence result was given for the case  $(r+1) \mid n$  and the finite field needs to be large enough. Later, in paper [9] and [2], the bound (1) was generalized to vector codes and nonlinear codes. The impact of field size on the minimum distance of LRCs was considered in [1]. The result provides an improved upper bound, but relies on a parameter related to another open problem in coding theory. In order to deal with multiple erasures in local repair, Prakash et al [11] proposed the locality  $(r, \delta)$  associating the coordinate with an inner-error-correcting code with length less than  $r + \delta - 1$  and minimum distance at least  $\delta$ . It is evident that the locality  $(r, \delta)$  degenerates into the locality  $r$  when  $\delta = 2$ . An upper bound was derived in [11] for codes with information locality  $(r, \delta)$  which coincides with the bound (1) at  $\delta = 2$ , and an explicit code attaining this bound was given for a specific value of the length  $n = \lceil \frac{k}{r} \rceil (r+1)$ .

For simplicity, the LRC that achieves the bound (1) with equality is usually called an optimal LRC. The first explicit optimal LRCs for the case  $(r+1) \mid n$  were constructed in [18] and [15] by using Reed-Solomon codes and Gabidulin codes respectively. Both constructions were built over a finite field of size exponential in the code length  $n$ . Moreover, it was proved in [15] that the construction also induces an optimal LRC when  $n \bmod (r+1) > k \bmod r > 0$ . Then in [19] for the same case  $(r+1) \mid n$  the authors constructed an optimal code over a finite field of size comparable to  $n$  by using specially designed polynomials. This construction can be extended to the case  $(r+1) \nmid n$  with the minimum distance  $d \geq n - k - \lceil \frac{k}{r} \rceil + 1$  which is at most one less than the upper bound defined in (1).

Recently, Song et al. [16] obtained more results about tightness of the bound (1). Specifically, they derived a new case where there are no optimal LRCs and two new cases where there exist optimal LRCs over sufficiently large fields, leaving only two cases in which tightness of the bound (1) is unknown. Another recent improvement was in [12] where Prakash et al. showed a new upper bound on the minimum distance for LRCs. This bound relies on a sequence of recursively defined parameters and is tighter than the bound (1). But no general constructions attaining this new bound was presented.

There are lots of other work devoted to the locality in the handling of multiple node failures, such as [13], [17], [19], [20] considering LRCs which permit parallel access of “hot data”, the papers [8], [20] studying LRCs with general local repair groups, and the work [12] which proposed sequential local repair. In a word, more and more research work have concerned about codes with the local repair property, especially those codes attaining the largest possible minimum distance.

### A. Our Contribution

Since the bound (1) is not tight for LRCs in many cases, the central problem in this work is determining the largest possible minimum distance of an  $[n, k]$  LRC.

Our first result is an integer programming based upper bound,

$$d \leq n - k + 1 - \eta,$$

where  $\eta = \max\{x : \Psi(x) - x < k\}$  and the function  $\Psi(x)$  relies on an integer programming problem defined below

$$\Psi(x) = \underset{\substack{s, t_1, \dots, t_s \\ a_1, \dots, a_s}}{\text{Max}} \underset{l, h_1, \dots, h_l}{\text{Min}} \left( xr + 1 - \sum_{i=1}^{l-1} (a_{h_i} - t_{h_i}) \right), \quad \forall 1 \leq x \leq \left\lceil \frac{n}{r+1} \right\rceil,$$

where the ‘Max’ is subject to

$$\begin{cases} t_1 + \dots + t_s = n_1; \\ a_1 + \dots + a_s = n_2; \\ a_i \geq t_i - 1, \forall i \in [s]; \\ s \geq 1; t_i \geq 1, \forall i \in [s], \end{cases}$$

and the ‘Min’ is subject to

$$t_{h_1} + \dots + t_{h_{l-1}} < x \leq t_{h_1} + \dots + t_{h_l}.$$

By solving the integer programming problem when  $n_1 > n_2$ , we get the second result of this paper: an explicit upper bound on the minimum distance (Theorem 14), where  $n_1 = \lceil \frac{n}{r+1} \rceil$  and  $n_2 = n_1(r+1) - n$ . This upper bound stands for all possible values of  $k$  while most previous results (e.g., [15], [16]) that depend on the value of  $k$  in addition to the parameters  $n$  and  $r$ , which means our bound sometimes covers wider parameter region. Additionally, in Section IV-B we show by comparisons that this explicit bound can give sharper description of the largest possible minimum distance than previous results (i.e. the results in [3], [12], [16]) in many cases.

The third result concerns the construction of LRCs. Specifically, when  $n_1 > n_2$ , we give an explicit construction (Construction 1) of the  $[n, k]$  LRC attaining the bound in Theorem 14 over the finite field  $\mathbb{F}_{2^m}$ , where  $m \geq n_1 r$ . Therefore, we have definitely determined the largest possible minimum distance for all  $[n, k]$  LRCs under the condition  $n_1 > n_2$ . Since the condition  $n \geq (r+1)^2$  implies  $n_1 > n_2$ , we have completely obtained the largest possible minimum distance for LRCs with  $r \leq \sqrt{n} - 1$ , which is of great significance in practical use.

### B. Related Work

In [21], the authors developed the framework of regenerating sets which determines the upper bound on the minimum distance for any LRC by computing a function related to the structure of local repair groups. The upper bound derived in this work can be viewed as an optimization based on this framework. A brief introduction of the framework and the motivation for optimization can be found in Section II.

### C. Organization

Section II introduces the framework of regenerating sets and shows the motivation of optimization. Section III derives an integer programming based upper bound on the minimum distance for LRCs. Then Section IV solves the integer programming problem for  $n_1 > n_2$ , and obtains an explicit upper bound. Section V presents an explicit construction attaining this bound. Finally, Section VI concludes the paper.

## II. REGENERATING SETS AND LOCALLY REPAIRABLE CODES

Let  $\mathcal{C}$  be an  $[n, k, d]_q$  linear code with generator matrix  $G = (\mathbf{g}_1, \dots, \mathbf{g}_n)$ , where  $\mathbf{g}_i \in \mathbb{F}_q^k$  for  $1 \leq i \leq n$ . Then the regenerating set introduced in [21] can be defined as follows.

**Definition 1.** For an  $[n, k, d]_q$  linear code  $\mathcal{C}$ , a regenerating set of the  $i$ th coordinate,  $1 \leq i \leq n$ , is a subset  $R \subseteq [n]$  such that  $i \in R$  and  $\mathbf{g}_i$  is an  $\mathbb{F}_q$ -linear combination of  $\{\mathbf{g}_j\}_{j \in R \setminus \{i\}}$ , where  $[n]$  denotes the set of integers  $\{1, 2, \dots, n\}$ .

The collection of all regenerating sets of the  $i$ th coordinate is denoted by  $\mathcal{R}_i$ . Furthermore, a sequence of regenerating sets  $R_1, R_2, \dots, R_m$ , where  $R_i \in \mathcal{R}_{l_i}$  and  $l_i \in [n]$  for  $1 \leq i \leq m$ , is said to have a *nontrivial union* if  $l_j \notin \cup_{i=1}^{j-1} R_i$  for  $1 \leq j \leq m$ .

For a linear code  $\mathcal{C}$ , define the function

$$\Phi(x) = \min\{|\cup_{i=1}^x R_i| : R_i \in \mathcal{R}_{l_i} \text{ and } R_1, \dots, R_x \text{ have a nontrivial union}\}. \quad (2)$$

In particular, it is assumed  $\Phi(0) = 0$ . Then it was proved that the minimum distance is closely related to the function  $\Phi(x)$ .

**Theorem 2** ([21]). For any  $[n, k, d]$  linear code,  $d \leq n - k + 1 - \rho$ , where  $\rho = \max\{x : \Phi(x) - x < k\}$ .

**Remark.** An explicit bound from Theorem 2 depends on computation of the function  $\Phi(x)$  which is determined by the specific generator matrix. Sometimes, partial information of the generator matrix may help get a precise estimate of  $\Phi(x)$  which in turn gives a tight bound for the minimum distance. An instance where Theorem 2 derives a tight bound is the square code proposed in [21]. In this paper, we aim to tighten the minimum distance bound for LRCs by estimating  $\Phi(x)$  and then optimizing the value. The following two subsections explain our motivations through examples.

### A. Estimate of $\Phi(x)$

First, we need to redefine the locality  $r$  by using the concept of regenerating sets.

**Definition 3.** For  $1 \leq i \leq n$ , the  $i$ th coordinate of an  $[n, k]$  code  $\mathcal{C}$  has locality  $r$  if there exists a regenerating set  $R \in \mathcal{R}_i$  with  $|R| \leq r + 1$ .

We refer to an  $[n, k]$  linear code of which each coordinate has locality  $r$  as a locally repairable code (LRC). Because  $r = 1$  implies repetition and for  $r \geq k$  MDS code possess the optimal distance, we assume  $1 < r < k$  throughout the paper. Moreover, because of the upper bound on the information rate of LRCs [19], we assume that  $\frac{k}{n} \leq \frac{r}{r+1}$  for any  $[n, k]$  LRC.

In [21] the authors estimated the function  $\Phi(x)$  for different kinds of locality and reproved the minimum distance bounds that had been given in previous literatures. For example, it proved  $\Phi(x) \leq (r+1)x$  for LRCs which induces the bound (1);  $\Phi(x) \leq r \lceil \frac{x}{\delta-1} \rceil + x$  for codes with locality  $(r, \delta)$  and derived the upper bound given in [11]; etc.

In this paper we focus on LRCs. The following example shows that when  $(r+1) \nmid n$  one can estimate  $\Phi(x)$  better than  $\Phi(x) \leq (r+1)x$  and thus can derive a tighter bound.

**Example 1.** Let  $\mathcal{C}$  be an  $[n, k, d]$  LRC with  $(r+1) \nmid n$ . We claim that  $\Phi(x) \leq x(r+1) - 1$  for  $x \geq 2$ .

First, the following algorithm generates a sequence of regenerating sets  $R_1, \dots, R_l$  that has a nontrivial union and  $\cup_{i=1}^l R_i = [n]$ .

```

1: Set  $i = 1$ 
2: while  $\cup_{j=1}^{i-1} R_j \subsetneq [n]$  do
3:   Pick  $i_0 \in [n] - \cup_{j=1}^{i-1} R_j$ 
4:   Choose  $R_i \in \mathcal{R}_{i_0}$  such that  $|R_i| = r + 1$ 
5:   Set  $i = i + 1$ 
6: end while

```

Because  $(r+1) \nmid n$  and  $|R_i| = r+1$  for  $1 \leq i \leq l$ , there exist  $i_1, i_2 \in [l]$  such that  $R_{i_1} \cap R_{i_2} \neq \emptyset$ . By the definition of  $\Phi(x)$ ,  $\Phi(x) \leq \min\{|\cup_{i \in I} R_i| : I \subset [l], |I| = x\}$ . Therefore,

$$\Phi(x) \leq \begin{cases} r+1, & \text{if } x = 1, \\ x(r+1) - 1, & \text{if } x \geq 2. \end{cases}$$

It follows that  $\rho \geq \lceil \frac{k+1}{r} \rceil - 1$ , and thus

$$d \leq n - k + 1 - \left( \left\lceil \frac{k+1}{r} \right\rceil - 1 \right). \quad (3)$$

Obviously, the bound (3) is tighter than the bound (1) for the case  $(r+1) \nmid n$ . Particularly, the difference occurs when  $r \mid k$  which also explains a known fact (see [3], [16]) that the bound (1) is unachievable when  $(r+1) \nmid n$  and  $r \mid k$ .

Later in Section III we will give a shaper estimate of  $\Phi(x)$  and derive a tighter bound for LRCs.

### B. Optimization of $\Phi(x)$

From Theorem 2 we observe that for a given LRC, its minimum distance  $d$  is upper bounded by  $n - k + 1 - \rho$ , where  $\rho$  depends on the function  $\Phi(x)$  which is determined by the code itself. Therefore, to upper bound  $d$  for all LRCs with parameters  $n, k, r$ , one needs to find the code which gives the minimum  $\rho$  or the maximum  $\Phi(x)$ . Actually, we find the structure of regenerating sets plays an important role in determining the function  $\Phi(x)$  which in turn influence the minimum distance.

**Example 2.** Consider LRCs with parameters  $n = 10, k = 5$  and  $r = 3$ . We construct two such LRCs which have different structure of regenerating sets.

The first code  $\mathcal{C}_1$  is constructed by using rank-metric codes [15]. Specifically, let

$$\{\alpha_1, \alpha_2, \alpha_3, \alpha_5, \alpha_6, \alpha_7, \alpha_9\} \subseteq \mathbb{F}_{2^7}$$

be a basis of  $\mathbb{F}_{2^7}$  over  $\mathbb{F}_2$  and let

$$\begin{cases} \alpha_4 = \alpha_1 + \alpha_2 + \alpha_3 \\ \alpha_8 = \alpha_5 + \alpha_6 + \alpha_7 \\ \alpha_{10} = \alpha_9. \end{cases}$$

The generator matrix of  $\mathcal{C}_1$  is  $G_1 = (\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_{10})$ , where  $\mathbf{g}_i = (\alpha_i, \alpha_i^2, \alpha_i^4, \alpha_i^8, \alpha_i^{16})^\tau$  for  $1 \leq i \leq 10$ .

It is easy to verify that  $\mathcal{C}_1$  is an LRC over  $\mathbb{F}_{2^7}$  and a sequence of its regenerating sets is

$$\{1, 2, 3, 4\}, \{5, 6, 7, 8\}, \{9, 10\}. \quad (4)$$

Therefore,  $\Phi(x) \leq 4x - 2$  for  $1 \leq x \leq 3$  and  $\rho \geq 2$ . By Theorem 2 we have  $d \leq n - k + 1 - \rho \leq 4$ . On the other hand, since any 7 columns of  $G_1$  has full rank, it implies  $d \geq n - 6 = 4$ . As a result,  $\mathcal{C}_1$  has minimum distance  $d = 4$ .

The second code  $\mathcal{C}_2$  is an  $[n = 10, k = 5]$  linear code over  $\mathbb{F}_{13}$  with generator matrix

$$G = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 5 & 5 & 11 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 3 & 7 & 10 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 10 & 10 & 7 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 6 & 3 & 9 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 10 & 9 & 6 \end{pmatrix}.$$

Observe that  $\mathcal{C}_2$  has locality  $r = 3$  and a sequence of its regenerating sets is

$$\{1, 2, 3, 4\}, \{1, 5, 6, 7\}, \{1, 8, 9, 10\}. \quad (5)$$

Furthermore, it can be verified that  $\Phi(1) = 4, \Phi(2) = 7$  and  $\Phi(3) = 10$ . Then  $\rho = 1$  and  $d \leq n - k + 1 - \rho \leq 5$  from Theorem 2. On the other hand, one can verify that  $\mathcal{C}_2$  has minimum distance  $d = 5$ .

From (4) and (5) we can see that  $\mathcal{C}_1$  and  $\mathcal{C}_2$  have different structure of regenerating sets. The former has pairwise disjoint regenerating sets while the latter has overlapped regenerating sets. This difference results in that the  $\Phi(x)$  of  $\mathcal{C}_1$  is no more than that of  $\mathcal{C}_2$ , therefore the latter code has a higher upper bound from Theorem 2.

### III. UPPER BOUNDS ON THE MINIMUM DISTANCE

Denote  $n_1 = \lceil \frac{n}{r+1} \rceil$  and  $n_2 = \lceil \frac{n}{r+1} \rceil (r+1) - n$ . It follows that  $n = n_1(r+1) - n_2$  and  $0 \leq n_2 < r+1$ . The integer programming based upper bound is derived in three steps as described in the following three subsections respectively.

#### A. From $\Phi(x)$ to a Set Cover Problem

First, for any  $[n, k]$  LRC, we convert the problem of estimating the  $\Phi(x)$  to a set cover problem (Lemma 5, Lemma 6). To begin with, we introduce the concept of an  $(r+1)$ -cover.

**Definition 4.** Let  $\mathcal{S} = \{S_1, \dots, S_t\}$  be a collection of subsets of  $[n]$ . We call  $\mathcal{S}$  an  $(r+1)$ -cover over  $[n]$  if the following conditions are satisfied:

- (1)  $|S_i| = r+1$  for  $1 \leq i \leq t$ ;
- (2)  $\cup_{i \in [t]} S_i = [n]$  and  $\cup_{i \in [t] \setminus \{j\}} S_i \neq [n]$  for any  $j \in [t]$ .

In the remainder of this paper we usually omit the phrase ‘over  $[n]$ ’ for an  $(r + 1)$ -cover when it is evident from the context.

**Lemma 5.** *For a given  $[n, k]$  locally repairable code  $\mathcal{C}$ , it induces an  $(r + 1)$ -cover  $\mathcal{S} = \{S_1, \dots, S_t\}$ ,  $t \geq n_1$ , satisfying*

$$\Phi(x) \leq \min_{\substack{J \subseteq [t] \\ |J|=x}} |\cup_{i \in J} S_i|$$

for  $1 \leq x \leq n_1$ , where  $\Phi(x)$  is defined as in (2).

*Proof:* By using the algorithm in Example 1, we get a sequence of regenerating sets  $R_1, \dots, R_l$  which has a nontrivial union. Then by deleting some  $R_i$ ’s which lie in the union of the remainders, we can finally get an  $(r + 1)$ -cover  $\{R_{i_1}, \dots, R_{i_t}\}$  as required by the lemma. ■

**Lemma 6.** *For any  $(r + 1)$ -cover  $\mathcal{S} = \{S_1, \dots, S_t\}$ ,  $t > n_1$ , there exists an  $(r + 1)$ -cover consisting of  $n_1$  subsets, denoted as  $\mathcal{T} = \{T_1, \dots, T_{n_1}\}$ , which satisfies for  $1 \leq x \leq n_1$ ,*

$$\min_{\substack{J \subseteq [t] \\ |J|=x}} |\cup_{i \in J} S_i| \leq \min_{\substack{I \subseteq [n_1] \\ |I|=x}} |\cup_{i \in I} T_i|.$$

*Proof:* Since  $t > n_1$ , set  $T_i = S_i$  initially for  $1 \leq i \leq n_1$ . Due to the condition (2) in Definition 4, it obviously has  $\cup_{i=1}^{n_1} T_i \subsetneq [n]$ . Then we recursively invoke the following Step 1 to Step 3 on the collection  $\mathcal{T} = \{T_1, \dots, T_{n_1}\}$  expanding  $\cup_{i=1}^{n_1} T_i$  by one element at each invocation until finally  $\cup_{i=1}^{n_1} T_i = [n]$ .

**Step 1.** Pick  $T_j \in \mathcal{T}$  such that  $T_j \cap (\cup_{T \in \mathcal{T} \setminus \{T_j\}} T) \neq \emptyset$ .

**Step 2.** Choose  $a \in T_j \cap (\cup_{T \in \mathcal{T} \setminus \{T_j\}} T)$  and  $b \in [n] - \cup_{i=1}^{n_1} T_i$ .

**Step 3.**  $T_j \leftarrow (T_j - \{a\}) \cup \{b\}$ .

Note that the subset  $T_j$  exists in Step 1 because  $\sum_{i=1}^{n_1} |T_i| = n_1(r + 1) \geq n > |\cup_{i=1}^{n_1} T_i|$ . After the three steps, only one element in  $T_j$  is replaced by an outside element and all other subsets remain unchanged. Therefore,  $\cup_{i=1}^{n_1} T_i$  is expanded by one element. Furthermore, the union size of any  $x$  subsets,  $1 \leq x \leq n_1$ , is unchanged or increased by 1. Therefore, for  $1 \leq x \leq n_1$ ,

$$\min_{\substack{J \subseteq [t] \\ |J|=x}} |\cup_{i \in J} S_i| \leq \min_{\substack{I \subseteq [n_1] \\ |I|=x}} |\cup_{i \in I} S_i| \leq \min_{\substack{I \subseteq [n_1] \\ |I|=x}} |\cup_{i \in I} T_i|.$$

Moreover, the condition  $\cup_{i \in [t] \setminus \{j\}} S_i \neq [n]$  for any  $j \in [t]$  implies that  $S_j \not\subseteq \cup_{i \in [t] \setminus \{j\}} S_i$  for any  $j \in [t]$ . It is easy to verify that the property  $T_j \not\subseteq \cup_{i \in [n_1] \setminus \{j\}} T_i$  for any  $j \in [n_1]$  still holds after an invocation of Step 1 to Step 3. Thus we finally get an  $(r + 1)$ -cover  $\mathcal{T}$  as the lemma requires. ■

By Lemma 5 and Lemma 6, we have transformed the problem of deriving an upper bound for  $\Phi(x)$  into the problem of estimating the set union size in an  $(r + 1)$ -cover consisting of  $n_1$  subsets. In the sequel, a further investigation into the  $(r + 1)$ -cover helps to finally derive an upper bound of  $\Phi(x)$ .

## B. From the Set Cover to an Integer Programming Problem

Then we transform the set cover problem into an integer programming problem (Lemma 11). The following definition comes from the concept of connectivity in graph theory.

**Definition 7.** *Let  $\mathcal{S} = \{S_1, \dots, S_t\}$  be a collection of nonempty subsets of  $[n]$ . We say  $\mathcal{S}$  is connected if for any nonempty subset  $I \subsetneq [t]$ , it has  $(\cup_{i \in I} S_i) \cap (\cup_{j \in [t] \setminus I} S_j) \neq \emptyset$ . Particularly, a collection containing only one subset, i.e.  $t = 1$ , is also called connected.*

**Remark.** In fact, a collection  $\mathcal{S}$  defines a graph  $G(V, E)$ , where each vertex  $v_i \in V$  corresponds to a subset  $S_i \in \mathcal{S}$  and there is an edge  $(v_i, v_j) \in E$  if and only if  $S_i \cap S_j \neq \emptyset$ . Thus a connected collection in Definition 7 actually corresponds to a connected graph.

**Proposition 8.** For a connected collection of subsets  $\mathcal{S} = \{S_1, \dots, S_t\}$ , there exists a permutation of  $[t]$ , say  $\{i_1, \dots, i_t\}$ , such that

$$S_{i_j} \cap (\cup_{h=1}^{j-1} S_{i_h}) \neq \emptyset, \quad 2 \leq j \leq t. \quad (6)$$

*Proof:* In fact,  $i_1, \dots, i_t$  can be determined by the following algorithm.

- 1: Pick  $i_1 \in [t]$
- 2: **for**  $h = 2$  **to**  $t$  **do**
- 3:   Pick  $i_h \in [t] - \{i_1, i_2, \dots, i_{h-1}\}$  such that  

$$S_{i_h} \cap (S_{i_1} \cup \dots \cup S_{i_{h-1}}) \neq \emptyset$$
- 4: **end for**

Note that the  $i_h$  at line 3 exists because the collection  $\mathcal{S}$  is connected. ■

**Corollary 9.** For a connected collection of subsets  $\mathcal{S} = \{S_1, \dots, S_t\}$ , define an integer  $a = \sum_{i=1}^t |S_i| - |\cup_{i=1}^t S_i|$ , then  $a \geq t - 1$ .

*Proof:* By Proposition 8, we can assume without loss of generality that  $\mathcal{S}$  satisfies the condition (6), i.e.,  $S_j \cap (\cup_{h=1}^{j-1} S_h) \neq \emptyset$  for all  $2 \leq j \leq t$ . Since

$$\begin{aligned} |\cup_{i=1}^t S_i| &= |S_t| - |S_t \cap (\cup_{i=1}^{t-1} S_i)| + |\cup_{i=1}^{t-1} S_i| \\ &= |S_t| - |S_t \cap (\cup_{i=1}^{t-1} S_i)| + |S_{t-1}| - |S_{t-1} \cap (\cup_{i=1}^{t-2} S_i)| + |\cup_{i=1}^{t-2} S_i| \\ &= \sum_{i=1}^t |S_i| - \sum_{i=2}^t |S_i \cap (\cup_{j=1}^{i-1} S_j)|, \end{aligned}$$

We have  $a = \sum_{i=2}^t |S_i \cap (\cup_{j=1}^{i-1} S_j)| \geq t - 1$ . ■

**Remark.** In the following, we introduce a set of integers to characterize the structure of an  $(r+1)$ -cover. First, for an  $(r+1)$ -cover  $\mathcal{S} = \{S_1, \dots, S_{n_1}\}$ , we determine a partition of  $[n_1]$ , say  $[n_1] = I_1 \cup \dots \cup I_s$ , such that

- (1) for  $1 \leq i \leq s$ , the induced collection  $\mathcal{S}_{I_i} = \{S_j \mid j \in I_i\}$  is connected; and
- (2) for  $1 \leq i < j \leq s$ ,  $(\cup_{h \in I_i} S_h) \cap (\cup_{h \in I_j} S_h) = \emptyset$ .

In other words, this partition of a collection  $\mathcal{S}$  actually corresponds to splitting the graph  $G(V, E)$  into connected components, where the graph  $G(V, E)$  is determined as in the remark after Definition 7. Then for  $1 \leq i \leq s$ , define integers  $t_i = |I_i|$  and  $a_i = \sum_{j \in I_i} |S_j| - |\cup_{j \in I_i} S_j|$ .

It is easy to derive the following lemma.

**Lemma 10.** For an  $(r+1)$ -cover  $\mathcal{S} = \{S_1, \dots, S_{n_1}\}$ , define integers  $s, t_1, \dots, t_s, a_1, \dots, a_s$  as in the above remark. Then the following conditions must hold:

$$\begin{cases} t_1 + \dots + t_s = n_1; \\ a_1 + \dots + a_s = n_2; \\ a_i \geq t_i - 1, \forall 1 \leq i \leq s; \\ s \geq 1; t_i \geq 1, \forall 1 \leq i \leq s. \end{cases} \quad (7)$$

*Proof:* By using the notations in the remark,  $I_1 \cup \dots \cup I_s$  is a partition of  $[n_1]$ , therefore  $a_1 + \dots + a_s = \sum_{i=1}^s (\sum_{j \in I_i} |S_j| - |\cup_{j \in I_i} S_j|) = \sum_{i=1}^{n_1} |S_i| - |\cup_{i=1}^{n_1} S_i| = n_1(r+1) - n = n_2$ . The other conditions come from Corollary 9 and the remark. ■

**Lemma 11.** For any  $(r + 1)$ -cover  $\mathcal{S} = \{S_1, \dots, S_{n_1}\}$ , define integers  $s, t_1, \dots, t_s, a_1, \dots, a_s$  as before, then for  $1 \leq x \leq n_1$ , it holds

$$\underset{\substack{I \subseteq [n_1] \\ |I|=x}}{\text{Min}} |\cup_{i \in I} S_i| \leq \underset{l, h_1, \dots, h_l}{\text{Min}} (xr + 1 - \sum_{i=1}^{l-1} (a_{h_i} - t_{h_i})),$$

where the ‘Min’ on the right side is subject to all integers  $l, h_1, \dots, h_l$  satisfying

$$t_{h_1} + \dots + t_{h_{l-1}} < x \leq t_{h_1} + \dots + t_{h_l}. \quad (8)$$

*Proof:* Suppose  $l$  and  $h_1, \dots, h_l$  are integers satisfying (8). Then there exists  $J \subseteq I_{h_l}$  such that  $|J| = x - (t_{h_1} + \dots + t_{h_{l-1}})$  and the collection  $\mathcal{S}_J$  is connected. It follows that

$$\begin{aligned} \underset{\substack{I \subseteq [n_1] \\ |I|=x}}{\text{Min}} |\cup_{i \in I} S_i| &\leq \sum_{j=1}^{l-1} |\cup_{i \in I_{h_j}} S_i| + |\cup_{i \in J} S_i| \\ &= \sum_{j=1}^{l-1} (\sum_{i \in I_{h_j}} |S_i| - a_{h_j}) + |\cup_{i \in J} S_i| \\ &\stackrel{(a)}{\leq} \sum_{j=1}^{l-1} (\sum_{i \in I_{h_j}} |S_i| - a_{h_j}) + \sum_{i \in J} |S_i| - (|J| - 1) \\ &= \sum_{j=1}^{l-1} (t_{h_j}(r + 1) - a_{h_j}) + |J|(r + 1) - (|J| - 1) \\ &\stackrel{(b)}{=} xr + 1 - \sum_{j=1}^{l-1} (a_{h_j} - t_{h_j}), \end{aligned}$$

where (a) is from Corollary 9 and (b) is from the equality that  $|J| = x - (t_{h_1} + \dots + t_{h_{l-1}})$ . ■

### C. An Integer Programming Based Bound

In this subsection, we derive an integer programming based bound on the minimum distance of any LRC (Theorem 12). Define

$$\Psi(x) = \underset{\substack{s, t_1, \dots, t_s \\ a_1, \dots, a_s}}{\text{Max}} \underset{l, h_1, \dots, h_l}{\text{Min}} (xr + 1 - \sum_{i=1}^{l-1} (a_{h_i} - t_{h_i})), \quad \forall 1 \leq x \leq n_1, \quad (9)$$

where the ‘Max’ is subject to (7) and the ‘Min’ is subject to (8). Then the value of  $\Psi(x)$  is determined only by integers  $n_1$  and  $n_2$ , or equivalently, by  $n$  and  $r$ .

**Theorem 12.** For any  $[n, k, d]$  LRC, it holds  $\Phi(x) \leq \Psi(x)$  for  $1 \leq x \leq n_1$ , and

$$d \leq n - k + 1 - \eta, \quad (10)$$

where  $\eta = \max\{x : \Psi(x) - x < k\}$ .

*Proof:* First, we show that  $\Phi(x) \leq \Psi(x)$ ,  $\forall 1 \leq x \leq n_1$ . By Lemma 5 and Lemma 6, there exists an  $(r + 1)$ -cover  $\mathcal{T}$  consisting of  $n_1$  subsets  $\{T_1, \dots, T_{n_1}\}$  such that

$$\Phi(x) \leq \underset{\substack{J \subseteq [n_1] \\ |J|=x}}{\text{Min}} |\cup_{j \in J} T_j|, \quad \forall 0 \leq x \leq n_1.$$



Define integers  $s, t_1, \dots, t_s, a_1, \dots, a_s$  as in the remark after Corollary 9. By Lemma 11 we have

$$\Phi(x) \leq \text{Min}_{l, h_1, \dots, h_l} (xr + 1 - \sum_{i=1}^{l-1} (a_{h_i} - t_{h_i})), \forall 1 \leq x \leq n_1,$$

where the minimum is subject to (8). Then it follows from Lemma 10 that  $\Phi(x) \leq \Psi(x)$  for  $1 \leq x \leq n_1$ . Therefore  $k > \Psi(\eta) - \eta \geq \Phi(\eta) - \eta$ . We have  $\eta \leq \rho$ , and then by Theorem 2, the bound (10) is obtained. ■

**Remark.** *Difference between the bound (10) and the bound in Theorem 2.* The two bounds are of the same form except that the former is determined by  $\eta$  and the function  $\Psi(x)$  while the latter is determined by  $\rho$  and the function  $\Phi(x)$ . But  $\Psi(x)$  is defined for all integers  $n$  and  $r$  while  $\Phi(x)$  is defined with respect to specific regenerating set structure. In other words, given parameters  $n$  and  $r$ , the bound (10) definitely provide an upper bound for any LRC with the parameters  $n$  and  $r$ , but Theorem 2 cannot give a specific bound due to the lack of information about regenerating set structure. Nevertheless, no efficient algorithm has been established for solving the integer programming problem involved in the bound (10). But we can solve it by exhaustive search for small  $n$  and  $r$  as in the example below. Furthermore, we can determine the solution for a wide class of the values of  $n$  and  $r$  which plays an important role in practical use. The details are in the next section.

**Example 3.** Suppose  $n = 13, r = 3$ , then  $n_1 = 4$  and  $n_2 = 3$ . Because of the assumption  $1 < r < k$  and the upper bound on the information rate of LRCs, i.e.  $\frac{k}{n} \leq \frac{r}{r+1}$ , we consider  $4 \leq k \leq 9$ .

First, compute the value of  $\Psi(x)$  for  $1 \leq x \leq 4$ . Observe that, up to permutation, all possible integers  $s$  and  $\{a_i, t_i\}_{i \in [s]}$  satisfying (7) are

$s = 1$	$t_1$	$a_1$	$s = 2$	$(t_1, t_2)$	$(a_1, a_2)$
	4	3		(1, 3)	(0, 3)
$s = 3$	$(t_1, t_2, t_3)$	$(a_1, a_2, a_3)$	$s = 4$	(1, 3)	(1, 2)
	(1, 1, 2)	(0, 0, 3)		(2, 2)	(1, 2)
	(1, 1, 2)	(0, 1, 2)		$(t_1, t_2, t_3, t_4)$	$(a_1, a_2, a_3, a_4)$
	(1, 1, 2)	(0, 2, 1)		(1, 1, 1, 1)	(0, 0, 0, 3)
	(1, 1, 2)	(1, 1, 1)		(1, 1, 1, 1)	(0, 0, 1, 2)
				(1, 1, 1, 1)	(0, 1, 1, 1)

Then by an exhaustive search, we get  $\Psi(1) = 4, \Psi(2) = 7, \Psi(3) = 10, \Psi(4) = 13$ . For simplicity, we can write  $\Psi(x) = 3x + 1$  for  $1 \leq x \leq 4$ .

Therefore we have  $\eta = \max\{x : \Psi(x) - x < k\} = \max\{x : 2x + 1 < k\} = \lceil \frac{k-3}{2} \rceil$  for  $4 \leq k \leq 9$ . Thus by Theorem 12,

$$d \leq n - k + 1 - \left\lceil \frac{k-3}{2} \right\rceil. \quad (11)$$

It gives an explicit upper bound. We compare it with the well known bound, i.e., the bound (1) given by Gopalan et al. As displayed in Fig. 1, the bound (11) goes through three points beneath the bound (1), i.e.  $k = 6, 9$  and  $8$ , where the former two points have been expected by the impossible condition  $(r+1) \nmid n$  and  $r \mid k$  (see Example 1) but the point  $k = 8$  is a new impossible result (not included in the impossible results in [16]).

#### IV. EXPLICIT BOUND FOR THE CASE $n_1 > n_2$

In this section, for a wide class of parameters, i.e.  $n_1 > n_2$ , we solve the integer programming problem involved in Theorem 12, and then derive an explicit upper bound for all LRCs satisfying  $n_1 > n_2$ . Since the condition  $n_1 > n_2$  can be viewed as a result of  $r \leq \sqrt{n} - 1$  which is a natural constraint for LRCs to be used in practice, the explicit bound we obtain here is sufficient to cover most practical use. In the second part of this section we make comparisons with all previously known results to show the improvements of our explicit bound. Actually, in Section V we will show this bound is tight for the case  $n_1 > n_2$ .

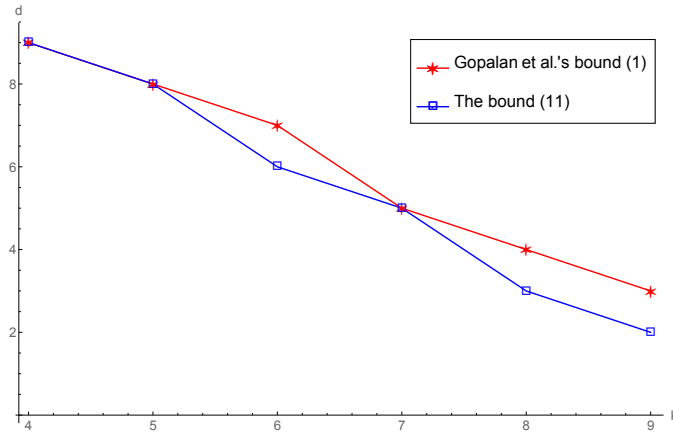


Fig. 1. Comparison of the two bounds for  $n = 13, r = 3$ .

#### A. Bound from Solution of the Integer Programming Problem

First, Proposition 13 determines the value of the function  $\Psi(x)$  under the condition  $n_1 > n_2$ . Then Theorem 14 derives an explicit upper bound accordingly.

Denote  $\mu = n_1 - n_2$  and let  $\lambda, \nu$  be integers such that  $n_1 = \lambda\mu + \nu$  and  $0 \leq \nu < \mu$ .

**Proposition 13.** For  $1 \leq x \leq n_1$ ,

$$\Psi(x) = xr + \max\left\{\left\lceil \frac{x}{\lambda + 1} \right\rceil, \left\lceil \frac{x - \nu}{\lambda} \right\rceil\right\}.$$

*Proof:* The proof is given in Appendix A. ■

**Theorem 14.** For any  $[n, k, d]$  LRC with  $n_1 > n_2$ , where  $n_1 = \lceil \frac{n}{r+1} \rceil$  and  $n_2 = n_1(r+1) - n$ , it holds

$$d \leq n - k + 1 - \tilde{\eta}, \quad (12)$$

where  $\tilde{\eta} = \min\left\{\left\lceil \frac{(\lambda+1)(k-1)+1}{(\lambda+1)(r-1)+1} \right\rceil, \left\lceil \frac{\lambda(k-1)+\nu+1}{\lambda(r-1)+1} \right\rceil\right\} - 1$ .

*Proof:* We prove this by showing  $\tilde{\eta} = \eta$ , where  $\eta$  is defined in Theorem 12. Specifically,

$$\begin{aligned} \eta &= \max\{x : \Psi(x) - x < k\} \\ &= \max\{x : x(r-1) + \max\left\{\left\lceil \frac{x}{\lambda+1} \right\rceil, \left\lceil \frac{x-\nu}{\lambda} \right\rceil\right\} < k\} \\ &= \max\{x : x(r-1) + \frac{x}{\lambda+1} \leq k-1 \text{ and } x(r-1) + \frac{x-\nu}{\lambda} \leq k-1\} \\ &= \max\{x : x \leq \frac{(\lambda+1)(k-1)}{(\lambda+1)(r-1)+1} \text{ and } x \leq \frac{\lambda(k-1)+\nu}{\lambda(r-1)+1}\}. \end{aligned}$$

Thus we have  $\eta = \min\left\{\left\lceil \frac{(\lambda+1)(k-1)+1}{(\lambda+1)(r-1)+1} \right\rceil, \left\lceil \frac{\lambda(k-1)+\nu+1}{\lambda(r-1)+1} \right\rceil\right\} - 1 = \tilde{\eta}$ , and the statement follows directly from Theorem 12. ■

**Example 4.** Let  $\mathcal{C}$  be an  $[n, k]$  LRC with  $(r+1) \mid n$ . We have  $n_1 = \frac{n}{r+1}, n_2 = 0$ , and therefore  $\mu = n_1, \nu = 0, \lambda = 1$ . Then it follows from Theorem 14 that  $\tilde{\eta} = \min\left\{\left\lceil \frac{2k-1}{2r-1} \right\rceil, \left\lceil \frac{k}{r} \right\rceil\right\} - 1 = \left\lceil \frac{k}{r} \right\rceil - 1$  and

$$d \leq n - k + 1 - \left(\left\lceil \frac{k}{r} \right\rceil - 1\right),$$

which coincides with the bound (1).

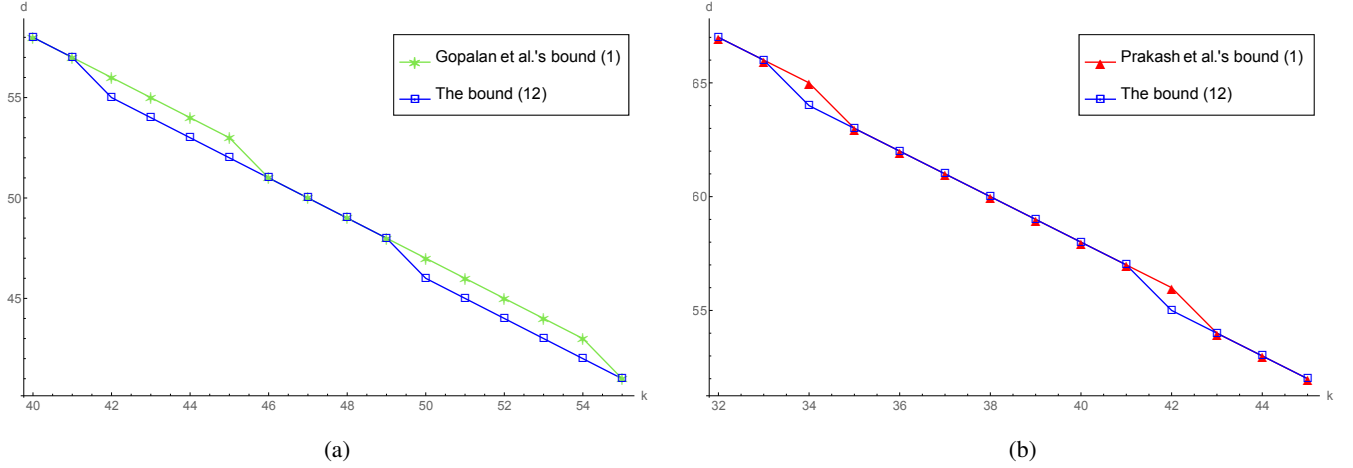


Fig. 2. Comparison of the three bounds for  $n = 101, r = 9$ .

### B. Improvements of the Bound

Since the bound (12) in Theorem 14 holds for  $n_1 > n_2$ , all the comparisons we make below are under the condition  $n_1 > n_2$ .

1) *Comparison with Gopalan et al.'s Bound:* The bound (1) given by Gopalan et al. [3] is the first upper bound on the minimum distance of LRCs. It states

$$d \leq n - k + 1 - \left( \left\lceil \frac{k}{r} \right\rceil - 1 \right).$$

Because  $n_1 > n_2$ , it follows  $\lambda \geq 1$  and  $\nu \geq 0$ . Then along with the assumption  $1 < r < k$ , a detailed calculation shows that  $\tilde{\eta} \geq \left\lceil \frac{k}{r} \right\rceil - 1$ . Therefore, the bound (12) generally provides a tighter upper bound than the bound (1). Actually, the former bound is strictly tighter than the latter at many points. The left graph of Fig. 2 gives a comparison of the two bounds for  $n = 101, r = 9$ .

2) *Comparison with Prakash et al.'s Bound:* Recently, Prakash et al. [12] derived an improved upper bound on the minimum distance, i.e.,

$$d \leq n - k + 1 - l, \quad (13)$$

where  $l$  is the unique integer satisfying  $e_l < k + l < e_{l+1}$  and  $\{e_m\}_{m \in [n_1]}$  is defined recursively as below,

$$e_{n_1} = n \quad \text{and} \quad e_{m-1} = e_m - \left\lceil \frac{2e_m}{m} \right\rceil + (r + 1) \quad \text{for } 2 \leq m \leq n_1.$$

It was proved in [12] that the bound (13) improves the bound (1). We claim that the bound (12) in Theorem 14 further improves the bound (13). Generally, observe that  $\eta = \max\{x : \Psi(x) - x < k\}$  and the definition of  $l$  is equivalent to  $l = \max\{m | e_m - m < k\}$ . Then the claim follows from the fact that

$$e_m \geq \Psi(m), \quad \forall 1 \leq m \leq n_1. \quad (14)$$

We prove (14) by induction on  $m$ .

First for  $m = n_1$ ,  $\Psi(n_1) = n_1 r + \mu = n = e_{n_1}$ . Then suppose the argument holds for  $m + 1$ , i.e.,

$e_{m+1} \geq \Psi(m+1)$ , where  $m < n_1$ . Thus

$$\begin{aligned}
e_m &= e_{m+1} - \left\lfloor \frac{2e_{m+1}}{m+1} \right\rfloor + (r+1) \\
&= \left\lfloor \frac{m-1}{m+1} e_{m+1} \right\rfloor + (r+1) \\
&\geq \left\lfloor \frac{m-1}{m+1} \Psi(m+1) \right\rfloor + (r+1) \\
&= \left\lfloor \frac{m-1}{m+1} ((m+1)r + \max\left\{ \left\lfloor \frac{m+1}{\lambda+1} \right\rfloor, \left\lfloor \frac{m+1-\nu}{\lambda} \right\rfloor \right\}) \right\rfloor + (r+1) \\
&= mr + 1 + \left\lfloor \frac{m-1}{m+1} \max\left\{ \left\lfloor \frac{m+1}{\lambda+1} \right\rfloor, \left\lfloor \frac{m+1-\nu}{\lambda} \right\rfloor \right\} \right\rfloor \\
&\geq mr + 1 + \left\lfloor \max\left\{ \frac{m-1}{\lambda+1}, \frac{m-1-\nu}{\lambda} \right\} \right\rfloor \\
&= mr + 1 + \max\left\{ \left\lfloor \frac{m}{\lambda+1} \right\rfloor - 1, \left\lfloor \frac{m-\nu}{\lambda} \right\rfloor - 1 \right\} \\
&= \Psi(m).
\end{aligned}$$

The above proof shows that the bound (12) cannot go upon the bound (13). A detailed calculation with specific values of  $n, k, r$  shows the former bound does go beneath the latter bound at some points. As an illustration, the right graph in Fig. 2 plots the two bounds for  $n = 101, r = 9$ .

3) *Comparing with the Results of Song et al* : In [16], Song et al. derived some conditions under which there exists *no* LRC attaining the bound (1), and also proved the existence of LRCs attaining the bound (1) under some conditions. However, they left some scope of parameters under which it was unknown whether there exist LRCs attaining the bound (1).

In Section V of this paper, we will give an explicit construction of  $[n, k]$  LRCs for  $n_1 > n_2$ , attaining the bound (12) in Theorem 14. Therefore our bound (12) completely describes the largest possible minimum distance for LRCs with  $n_1 > n_2$ .

Fig. 3 illustrates the corresponding results for  $n = 50, 10 \leq k \leq 17$  and  $2 \leq r \leq 9$ . In the tables ‘Y’ means there exist LRCs attaining the bound (1), ‘N’ means there is *no* LRC attaining the bound (1), and a blank means it is unknown whether there exist LRCs attaining the bound (1).

Results of Song et al.									Results of this paper								
$r \backslash k$	10	11	12	13	14	15	16	17	$r \backslash k$	10	11	12	13	14	15	16	17
2	N		N		N		N		2	N	Y	N	Y	N	Y	N	Y
3	Y	N	N	Y	N	N		N	3	Y	N	N	Y	N	N	Y	N
4	Y	Y	Y	Y	Y	Y	Y	Y	4	Y	Y	Y	Y	Y	Y	Y	Y
5	N	Y	Y	Y	Y	N	Y	Y	5	N	Y	Y	Y	Y	N	Y	Y
6	Y		N	Y	Y				6	Y	Y	N	Y	Y	Y	Y	N
7	Y	Y	Y		N	Y	Y	Y	7	Y	Y	Y	Y	N	Y	Y	Y
8	Y	Y	Y	Y	Y		N	Y	8	Y	Y	Y	Y	Y	Y	N	Y
9	Y	Y	Y	Y	Y	Y	Y	Y	9	Y	Y	Y	Y	Y	Y	Y	Y

Fig. 3. A comparison of Song et al.’s results and our results for  $n = 50$ .

## V. CODE CONSTRUCTION WHEN $n_1 > n_2$

In this section, we present an explicit construction of LRCs attaining the bound (12). The construction is based on linearized polynomials. We start this section with some basic facts about linearized polynomials.

### A. The Linearized Polynomial

**Definition 15.** A polynomial of the form  $f(x) = \sum_{i=0}^t a_i x^{q^i}$  with coefficients  $a_i \in \mathbb{F}_{q^m}$  for  $0 \leq i \leq t$  and  $a_t \neq 0$  is called a linearized polynomial of  $q$ -degree  $t$  over the extension field  $\mathbb{F}_{q^m}$ .

A linearized polynomial  $f(x)$  can be viewed as an  $\mathbb{F}_q$ -linear transformation from  $\mathbb{F}_{q^m}$  to itself, i.e., for any  $c_1, c_2 \in \mathbb{F}_q$  and  $\omega_1, \omega_2 \in \mathbb{F}_{q^m}$ , it holds  $f(c_1\omega_1 + c_2\omega_2) = c_1f(\omega_1) + c_2f(\omega_2)$ . Furthermore, a standard result of finite fields states that,

**Proposition 16.** [6] A linearized polynomial  $f(x)$  of  $q$ -degree no more than  $t$  can be uniquely determined by the values of  $f(\omega_1), \dots, f(\omega_{t+1})$ , where  $\omega_1, \dots, \omega_{t+1}$  are  $t+1$  elements in  $\mathbb{F}_{q^m}$  that are linearly independent over  $\mathbb{F}_q$ .

### B. An Explicit Code Construction

In this subsection, we assume  $n_1 > n_2$  and construct an  $[n, k]$  LRC over  $\mathbb{F}_{q^m}$  attaining the bound (12) in Theorem 14, where  $\mathbb{F}_{q^m}$  is an extension field of  $\mathbb{F}_q$  with  $m \geq n_1 r$ . In a word, the codewords are obtained as evaluations of a linearized polynomial at  $n$  points in  $\mathbb{F}_{q^m}$ . Because of the property of linearized polynomials introduced in Proposition 16, the key point of the code construction is the selection of the  $n$  evaluation points such that the resulting code has the largest possible minimum distance. Denote the set of the  $n$  evaluation points by  $\Omega$ .

Since  $\mathbb{F}_{q^m}$  can be viewed as an  $\mathbb{F}_q$ -linear space of dimension  $m$ , by fixing a basis of  $\mathbb{F}_{q^m}$  over  $\mathbb{F}_q$ , the  $n$  elements in  $\Omega$  can be expressed as  $n$  vectors of length  $m$  over  $\mathbb{F}_q$ . These  $n$  vectors are determined through the following three steps. For simplicity, we can set  $q = 2$  and  $m = n_1 r$ , and the process below also works for other values of  $q$  and  $m$ .

*Step 1.* Let  $X = (\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_r)$  be the generator matrix of an  $[r+1, r]_2$  MDS code and let  $\mathbf{c} = (1, c_1, \dots, c_r)$  be one of its codeword, where  $\mathbf{x}_i \in \mathbb{F}_2^r$  for  $0 \leq i \leq r$ . For example, we can choose

$$X = \begin{pmatrix} 1 & 0 & \dots & 0 & 1 \\ 0 & 1 & \dots & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \dots & 1 & 1 \end{pmatrix} \text{ and } \mathbf{c} = (1, 0, \dots, 0, 1).$$

*Step 2.* Define vectors  $\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_{i,j} \in \mathbb{F}_2^{(\lambda+1)r}$  for  $1 \leq i \leq \lambda+1$  and  $1 \leq j \leq r$ , where

$$\boldsymbol{\alpha}_0 = \begin{pmatrix} \mathbf{x}_0 \\ \mathbf{x}_0 \\ \vdots \\ \mathbf{x}_0 \end{pmatrix} \text{ and } \boldsymbol{\alpha}_{i,j} = \begin{pmatrix} c_j \mathbf{x}_0 \\ \vdots \\ \mathbf{x}_j \\ \vdots \\ c_j \mathbf{x}_0 \end{pmatrix},$$

that is,  $\boldsymbol{\alpha}_0$  consists of  $(\lambda+1)$   $\mathbf{x}_0$ 's and  $\boldsymbol{\alpha}_{i,j}$  is defined by replacing the  $i$ -th  $c_j \mathbf{x}_0$  of  $c_j \boldsymbol{\alpha}_0$  with an  $\mathbf{x}_j$ . Similarly, define  $\boldsymbol{\beta}_0, \boldsymbol{\beta}_{i,j} \in \mathbb{F}_2^{\lambda r}$ ,  $1 \leq i \leq \lambda$  and  $1 \leq j \leq r$ , such that  $\boldsymbol{\beta}_0$  consists of  $\lambda$   $\mathbf{x}_0$ 's and  $\boldsymbol{\beta}_{i,j}$  is defined by replacing the  $i$ -th  $c_j \mathbf{x}_0$  of  $c_j \boldsymbol{\alpha}_0$  with an  $\mathbf{x}_j$ . For example, let  $r = 2$ ,  $\lambda = 2$  and  $X = (\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2)$ ,  $\mathbf{c} = (1, c_1, c_2)$ , then we have

$$(\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_{1,1}, \boldsymbol{\alpha}_{1,2}, \boldsymbol{\alpha}_{2,1}, \boldsymbol{\alpha}_{2,2}, \boldsymbol{\alpha}_{3,1}, \boldsymbol{\alpha}_{3,2}) = \begin{pmatrix} \mathbf{x}_0 & \mathbf{x}_1 & \mathbf{x}_2 & c_1 \mathbf{x}_0 & c_2 \mathbf{x}_0 & c_1 \mathbf{x}_0 & c_2 \mathbf{x}_0 \\ \mathbf{x}_0 & c_1 \mathbf{x}_0 & c_2 \mathbf{x}_0 & \mathbf{x}_1 & \mathbf{x}_2 & c_1 \mathbf{x}_0 & c_2 \mathbf{x}_0 \\ \mathbf{x}_0 & c_1 \mathbf{x}_0 & c_2 \mathbf{x}_0 & c_1 \mathbf{x}_0 & c_2 \mathbf{x}_0 & \mathbf{x}_1 & \mathbf{x}_2 \end{pmatrix}$$

and

$$(\beta_0, \beta_{1,1}, \beta_{1,2}, \beta_{2,1}, \beta_{2,2}) = \begin{pmatrix} \mathbf{x}_0 & \mathbf{x}_1 & \mathbf{x}_2 & c_1 \mathbf{x}_0 & c_2 \mathbf{x}_0 \\ \mathbf{x}_0 & c_1 \mathbf{x}_0 & c_2 \mathbf{x}_0 & \mathbf{x}_1 & \mathbf{x}_2 \end{pmatrix}.$$

The vectors  $\alpha_0, \alpha_{i,j}$  and  $\beta_0, \beta_{i,j}$  defined above have the following properties.

**Lemma 17.** Denote  $\mathcal{A}_i = \{\alpha_0, \alpha_{i,1}, \dots, \alpha_{i,r}\}$  for  $1 \leq i \leq \lambda + 1$ . Then we have

- (i) For  $1 \leq i \leq \lambda + 1$ , each vector contained in  $\mathcal{A}_i$  is an  $\mathbb{F}_q$ -linear combination of the other  $r$  vectors in  $\mathcal{A}_i$ .
- (ii) For any  $F \subseteq \cup_{i=1}^{\lambda+1} \mathcal{A}_i$  satisfying that there exists a vector  $\alpha \in F$  such that  $|(F - \{\alpha\}) \cap \mathcal{A}_i| \leq r - 1$  for  $1 \leq i \leq \lambda + 1$ , the vectors in  $F$  are  $\mathbb{F}_q$ -linearly independent.

Denote  $\mathcal{B}_j = \{\beta_0, \beta_{j,1}, \dots, \beta_{j,r}\}$  for  $1 \leq j \leq \lambda$ . Then the same statements also hold for  $\mathcal{B}_j$  for  $1 \leq j \leq \lambda$ .

*Proof:* The proof is given in Appendix B. ■

*Step 3.* Let  $A$  be the matrix consisting of the  $((\lambda + 1)r + 1)$  column vectors in  $\cup_{i=1}^{\lambda+1} \mathcal{A}_i$ , and let  $B$  be the matrix consisting of the  $(\lambda r + 1)$  column vectors in  $\cup_{i=1}^{\lambda} \mathcal{B}_i$ . Define a block diagonal matrix

$$W = \begin{pmatrix} A & & & & \\ & \ddots & & & \\ & & A & & \\ & & & B & \\ & & & & \ddots \\ & & & & & B \end{pmatrix}$$

which is composed of  $\nu$   $A$ 's and  $(\mu - \nu)$   $B$ 's on the diagonal and zeros elsewhere. Note that  $A$  has  $((\lambda + 1)r + 1)$  columns and  $B$  has  $(\lambda r + 1)$  columns, then  $W$  has  $((\lambda + 1)r + 1)\nu + (\lambda r + 1)(\mu - \nu) = n_1(r + 1) - n_2 = n$  columns. Similarly,  $W$  has  $\nu(\lambda + 1)r + (\mu - \nu)\lambda r = (\lambda\mu + \nu)r = n_1 r$  rows. Then the set of  $n$  vectors in  $\Omega$  are defined to be the  $n$  columns of  $W$ .

We give a graphical explanation of linear dependences among the  $n$  vectors. Refer to Fig. 4, each point actually corresponds to a vector. Then the left  $\nu$  trees each composed of  $\lambda + 1$  branches corresponds to the  $\nu$  blocks of  $A$  in  $W$ , and the right  $\mu - \nu$  trees each composed of  $\lambda$  branches corresponds to the  $\mu - \nu$  blocks of  $B$  in  $W$ . In more detail, the set  $\mathcal{A}_i$  for  $1 \leq i \leq \lambda + 1$  corresponds to a branch in the left trees and particularly the vector  $\alpha_0$  corresponds to the root point. The similar correspondence holds for  $\mathcal{B}_i$  and the branches in the right trees.

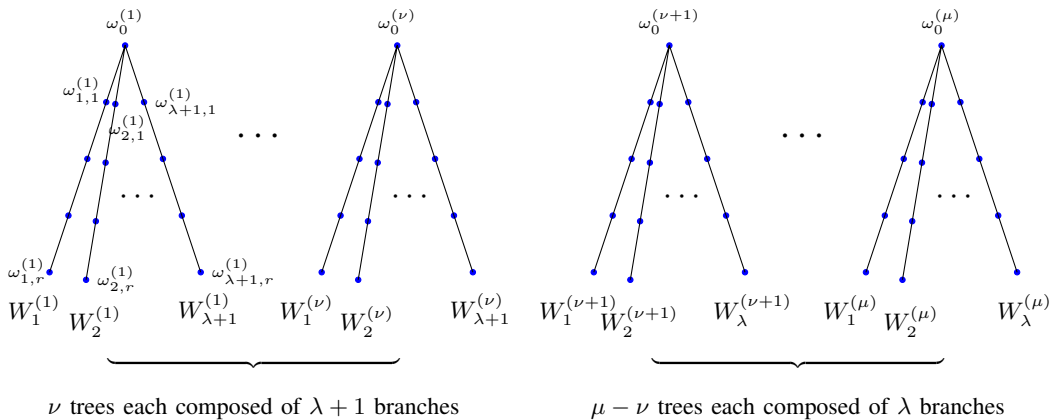


Fig. 4. The  $n$  points in  $\Omega$ .

For convenience, we denote the  $n$  points (or equivalently, the  $n$  vectors in  $\Omega$ ) by

$$\{\omega_0^{(l)}, \omega_{i,j}^{(l)} \mid l \in [\mu], i \in [\lambda+1], j \in [r]\}^1$$

where the superscript  $l$  denotes which tree it belongs to, the subscript  $i$  denotes which branch it lies in and  $j$  is the point index in that branch. Moreover, denote each branch by

$$W_i^{(l)} = \{\omega_0^{(l)}, \omega_{i,1}^{(l)}, \omega_{i,2}^{(l)}, \dots, \omega_{i,r}^{(l)}\} \text{ for } l \in [\mu] \text{ and } i \in [\lambda+1].$$

Then by Lemma 17 (i), each vector in  $W_i^{(l)}$  is an  $\mathbb{F}_q$ -linear combination of the other  $r$  vectors in  $W_i^{(l)}$ , and by the construction of the matrix  $W$ , the vectors in different trees are linearly independent.

**Construction 1.** Define an  $[n, k]$  linear code  $\mathcal{C}$  over  $\mathbb{F}_{q^m}$  as follows.

- Let  $\Omega \subseteq \mathbb{F}_{q^m}$  be a set of the  $n$  vectors defined above, i.e.,  $\Omega = \{\omega_0^{(l)}, \omega_{i,j}^{(l)} \mid l \in [\mu], i \in [\lambda+1], j \in [r]\}$ . Note that each vector is of length  $n_1 r = m$  over  $\mathbb{F}_q$  and thus can be viewed as an element in  $\mathbb{F}_{q^m}$ .
- $\mathcal{C}$  encodes a file  $(m_0, \dots, m_{k-1}) \in \mathbb{F}_{q^m}^k$  into  $(f(\omega))_{\omega \in \Omega} \in \mathbb{F}_{q^m}^n$ , where  $f(x) = \sum_{i=0}^{k-1} m_i x^{q^i}$ .

Denote the  $n$  coordinates of  $\mathcal{C}$  by the corresponding element in  $\Omega$ , then  $W_i^{(l)}$  is a regenerating set of each coordinate contained in  $W_i^{(l)}$ . Therefore,  $\mathcal{C}$  is an  $[n, k]$  LRC with locality  $r$ .

**Example 5.** We illustrate the construction through a specific example. Suppose  $n = 8, k = 4, r = 2$ , then it has  $n_1 = 3, n_2 = 1$  and  $\lambda = 1, \mu = 2, \nu = 1$ .

The construction is over the field  $\mathbb{F}_{2^6} = \mathbb{F}_2(\theta)$ , where  $\theta$  is a primitive element of  $\mathbb{F}_{2^6}$  with minimal polynomial  $x^6 + x^5 + 1$ . By fixing a basis  $\{1, \theta, \theta^2, \dots, \theta^5\}$ , the subset  $\Omega \subseteq \mathbb{F}_{2^6}$  is constructed as follows.

First, let

$$X = (\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2) = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \text{ and } \mathbf{c} = (1, c_1, c_2) = (1, 0, 1).$$

Then

$$\begin{aligned} A &= (\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_{1,1}, \boldsymbol{\alpha}_{1,2}, \boldsymbol{\alpha}_{2,1}, \boldsymbol{\alpha}_{2,2}) \\ &= \begin{pmatrix} \mathbf{x}_0 & \mathbf{x}_1 & \mathbf{x}_2 & c_1 \mathbf{x}_0 & c_2 \mathbf{x}_0 \\ \mathbf{x}_0 & c_1 \mathbf{x}_0 & c_2 \mathbf{x}_0 & \mathbf{x}_1 & \mathbf{x}_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \end{aligned}$$

$$\begin{aligned} \text{and } B &= (\boldsymbol{\beta}_0, \boldsymbol{\beta}_{1,1}, \boldsymbol{\beta}_{1,2}) \\ &= (\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2) = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}. \end{aligned}$$

Therefore

$$W = \begin{pmatrix} A & B \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix},$$

and thus

$$\begin{aligned} \omega_0^{(1)} &= 1 + \theta^2 & \omega_0^{(2)} &= \theta^4 \\ \omega_{1,1}^{(1)} &= \theta & \omega_{1,1}^{(2)} &= \theta^5 \\ \omega_{1,2}^{(1)} &= 1 + \theta + \theta^2 & \omega_{1,2}^{(2)} &= \theta^4 + \theta^5 \\ \omega_{2,1}^{(1)} &= \theta^3 & & \\ \omega_{2,2}^{(1)} &= 1 + \theta^2 + \theta^3 & & \end{aligned} \quad \text{and}$$

Fig. 5 gives a graphical illustration of the eight elements in  $\Omega$ .

<sup>1</sup> A precise description of the range of  $l$  and  $i, j$  is  $(l, i) \in ([\nu] \times [\lambda+1]) \cup ([\nu+1, \mu] \times [\lambda]), j \in [r]$ , where  $[\nu+1, \mu] = \{\nu+1, \nu+2, \dots, \mu\}$ .

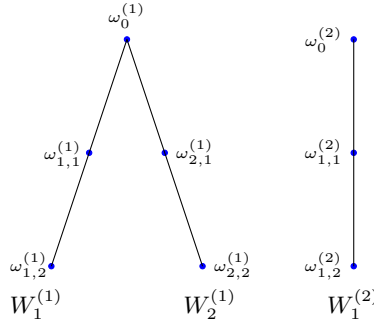


Fig. 5. The eight elements of  $\Omega$  for the  $[8, 4]$  code.

Then the  $[n = 8, k = 4]$  linear code  $\mathcal{C}$  encodes a file  $(m_0, m_1, m_2, m_3)$  into  $(f(\omega))_{\omega \in \Omega}$ , where  $f(x) = m_0x + m_1x^2 + m_2x^4 + m_3x^8$ .

A sequence of regenerating sets of the linear code  $\mathcal{C}$  is

$$\{\omega_0^{(1)}, \omega_{1,1}^{(1)}, \omega_{1,2}^{(1)}\}, \{\omega_0^{(1)}, \omega_{2,1}^{(1)}, \omega_{2,2}^{(1)}\}, \{\omega_0^{(2)}, \omega_{1,1}^{(2)}, \omega_{1,2}^{(2)}\},$$

and it is easy to see that  $\Phi(1) = 3, \Phi(2) = 5, \Phi(3) = 8$ , which coincides with the upper bound defined by  $\Psi(x)$  (see Proposition 13). Moreover, it can be verified that the minimum distance of  $\mathcal{C}$  is  $d = 3$ , which is optimal with respect to the bound (12) in Theorem 14. Actually, the following theorem states that the code  $\mathcal{C}$  in Construction 1 always attains the bound (12) in Theorem 14.

**Theorem 18.** *The  $[n, k]$  LRC  $\mathcal{C}$  obtained from Construction 1 has the minimum distance*

$$d = n - k + 1 - \tilde{\eta},$$

where  $\tilde{\eta} = \min\left\{\left\lceil \frac{(\lambda+1)(k-1)+1}{(\lambda+1)(r-1)+1} \right\rceil, \left\lceil \frac{\lambda(k-1)+\nu+1}{\lambda(r-1)+1} \right\rceil\right\} - 1$ .

*Proof:* First, we claim that for any  $V \subseteq \Omega$  with  $|V| = k + \tilde{\eta}$ , there exist subsets  $V_1, \dots, V_\mu \subseteq V$  such that the following two conditions are satisfied:

- (1)  $|\cup_{l=1}^\mu V_l| \geq k$ ;
- (2) For  $1 \leq l \leq \mu$ ,  $V_l \subseteq \cup_{i=1}^{\lambda+1} W_i^{(l)}$ , and there exists  $\omega_l \in V_l$  such that  $|(V_l - \{\omega_l\}) \cap W_i^{(l)}| \leq r - 1$  for all  $i \in [\lambda + 1]$ .

The proof of the claim is given in Lemma 21 of Appendix C.

From the claim and Lemma 17 (ii), we can deduce that, for  $1 \leq l \leq \mu$ , the elements in  $V_l$  are linearly independent over  $\mathbb{F}_q$ , and thus the elements in  $V_1 \cup V_2 \cup \dots \cup V_\mu$  are linearly independent over  $\mathbb{F}_q$ . Then by Proposition 16,  $\mathcal{C}$  can tolerate any  $n - (k + \tilde{\eta})$  erasures. Consequently, the minimum distance of  $\mathcal{C}$  satisfies  $d \geq n - k + 1 - \tilde{\eta}$ , and the equality actually holds because of Theorem 14. ■

### C. Influence of the Regenerating Set Structure

As we have stated in Example 2 and earlier sections, the structure of regenerating sets can influence the value of the function  $\Phi(x)$  which in turn influence the value of the minimum distance. In this section, we will check the regenerating set structure of the code  $\mathcal{C}$  in Construction 1 to support its attaining the optimal minimum distance, and also make a comparison with some previously constructed codes.

In Fig. 4 it gives a graphical description of the regenerating sets for  $\mathcal{C}$ , while each line (or a branch, i.e.  $W_i^{(l)}$ ) represents a regenerating set. Consider the collection of regenerating sets  $\{W_i^{(l)}\}_{l \in [\mu], i \in [\lambda+1]}$ . It has a nontrivial union with respect to any order they are arranged in.

In fact, it is easy to see that for the code  $\mathcal{C}$ ,

$$\Phi(x) = \min_{\substack{\mathcal{I} \subseteq \{W_i^{(l)}\}_{l \in [\mu], i \in [\lambda+1]} \\ |\mathcal{I}| = x}} |\cup_{V \in \mathcal{I}} V|.$$



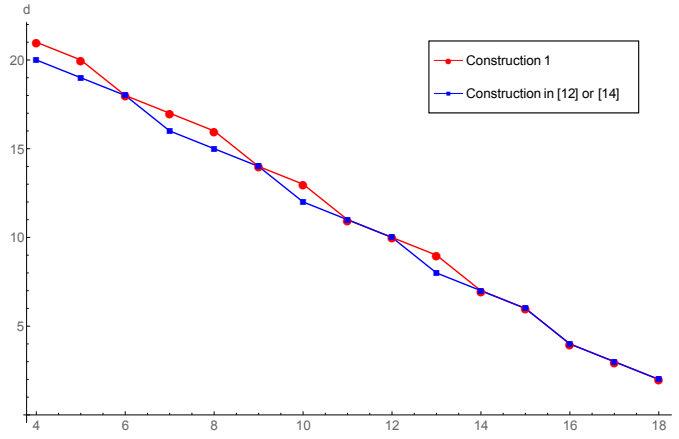


Fig. 6. A comparison of the two LRCs for  $n = 25, r = 3$

We can count from Fig. 4 that

$$\begin{aligned} \min_{\substack{\mathcal{I} \subseteq \{W_i^{(l)}\}_{l,i} \\ |\mathcal{I}|=x}} |\cup_{V \in \mathcal{I}} V| &= \begin{cases} xr + \left\lceil \frac{x}{\lambda+1} \right\rceil, & \text{if } x \leq \nu(\lambda+1) \\ xr + \nu + \left\lceil \frac{x - \nu(\lambda+1)}{\lambda} \right\rceil, & \text{if } x > \nu(\lambda+1) \end{cases} \\ &= xr + \max\left\{ \left\lceil \frac{x}{\lambda+1} \right\rceil, \left\lceil \frac{x - \nu}{\lambda} \right\rceil \right\}. \end{aligned}$$

Therefore, the  $\Phi(x)$  of  $\mathcal{C}$  satisfies

$$\Phi(x) = xr + \max\left\{ \left\lceil \frac{x}{\lambda+1} \right\rceil, \left\lceil \frac{x - \nu}{\lambda} \right\rceil \right\},$$

which attains the upper bound defined by  $\Psi(x)$  (see Theorem 12 and Proposition 13). That is,  $\mathcal{C}$  achieves the maximum value of  $\Phi(x)$  among all the LRCs with  $n_1 > n_2$ , which can be regarded as a support of the code  $\mathcal{C}$  attaining the optimal minimum distance.

On the other hand, we will see some previously constructed codes have smaller minimum distance due to their regenerating set structure. The code presented by Silberstein et al. in [15] and that proposed by Tamo et al. in [19] are both of pairwise disjoint regenerating sets. Namely, partition the set  $[n]$  into  $n_1$  subsets  $I_1, I_2, \dots, I_{n_1}$  such that  $|I_j| = r+1$  for  $1 \leq j \leq n_1 - 1$  and  $|I_{n_1}| = r+1 - n_2$ , then  $I_1, I_2, \dots, I_{n_1}$  form a sequence of regenerating sets that has a nontrivial union.

Clearly, the  $\Phi(x)$  satisfies

$$\Phi(x) \leq (r+1)x - n_2, \quad \forall 1 \leq x \leq n_1.$$

Then by Theorem 2,  $\rho = \max\{x : \Phi(x) - x < k\} \geq \left\lceil \frac{k+n_2}{r} \right\rceil - 1$ , and the minimum distance satisfies

$$d \leq n - k + 1 - \left( \left\lceil \frac{k+n_2}{r} \right\rceil - 1 \right).$$

Thus it cannot attain the bound (1) when  $\left\lceil \frac{k+n_2}{r} \right\rceil > \left\lceil \frac{k}{r} \right\rceil$ , i.e.,  $k \bmod r \geq n \bmod (r+1) > 0$ . In fact, the minimum distance sometimes goes beneath the bound (12) of Theorem 14, that is, the optimal minimum distance cannot be attained under this kind of regenerating set structure. Fig. 6 gives a comparison between the minimum distance of  $\mathcal{C}$  and that of the codes in [15], [19] for  $n = 25$  and  $r = 3$ .

## VI. CONCLUSIONS

In this paper we carry out an in-depth study of the two problems: what is the largest possible minimum distance for an  $[n, k]$  LRC? How to construct an  $[n, k]$  LRC with the largest possible minimum distance? For the first problem, we derive an integer programming based upper bound on the minimum distance for LRCs, and then give an explicit bound by solving the integer programming problem. The explicit bound applies all LRCs satisfying  $n_1 > n_2$ . For the second problem, we present a construction of linear LRCs that attains the explicit bound for  $n_1 > n_2$ . Therefore, we have completely solved the two problems under the condition  $n_1 > n_2$ . However, for  $n_1 \leq n_2$  the two problems remain unsolved in many cases.

## REFERENCES

- [1] V. Cadambe and A. Mazumdar, "An upper bound on the size of locally recoverable codes," *IEEE Int. Symp. Netw. Coding (NetCod)*, Calgary, 2013, pp. 1–5.
- [2] M. Forbes and S. Yekhanin, "On the locality of codeword symbols in non-linear codes," *arXiv preprint arXiv:1303.3921*, 2013.
- [3] P. Gopalan, C. Huang, H. Simitci, and S. Yekhanin, "On the locality of codeword symbols," *IEEE Trans. on Inform. Theory*, vol. 58, pp. 6925–6934, Nov. 2012.
- [4] C. Huang, M. Chen, and J. Li, "Pyramid codes: Flexible schemes to trade space for access efficiency in reliable data storage systems," in *Proc. 6th IEEE Int. Symp. Netw. Comput. Appl.*, Cambridge, 2007, pp. 79C86.
- [5] C. Huang, H. Simitci, Y. Xu, A. Ogus, B. Calder, P. Gopalan, J. Li, and S. Yekhanin, "Erasure coding in Windows Azure Storage," presented at the USENIX Annu. Tech. Conf., Boston, MA, 2012.
- [6] R. Lidl, *Finite fields*, Cambridge University Press, 1997.
- [7] F. Oggier and A. Datta, "Self-repairing homomorphic codes for distributed storage systems," in *Proc. IEEE Infocom*, Shanghai, 2011, pp. 1215–1223.
- [8] L. Parnes-Juarez, H. D. L. Hollmann, and F. Oggier, "Locally repairable codes with multiple repair alternatives," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Istanbul, 2013, pp. 892–896.
- [9] D. S. Papailiopoulos and A. G. Dimakis, "Locally repairable codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Cambridge, 2012, pp. 2771–2775.
- [10] D. S. Papailiopoulos, J. Luo, A. G. Dimakis, C. Huang, , and J. Li, "Simple regenerating codes: network coding for cloud storage," in *Proc. IEEE Infocom*, Orlando, 2012, pp. 2801–2805.
- [11] N. Prakash, G. M. Kamath, V. Lalitha, and P. V. Kumar, "Optimal linear codes with a local-error-correction property," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Cambridge, 2012, pp. 2776–2780.
- [12] N. Prakash, V. Lalitha, and P. Kumar, "Codes with locality for two erasures," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Honolulu, 2014, pp. 1962–1966.
- [13] A. S. Rawat, D. S. Papailiopoulos, A. G. Dimakis, and S. Vishwanath, "Locality and availability in distributed storage," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Honolulu, 2014, pp. 681–685.
- [14] M. Sathiamoorthy, M. Asteris, D. Papailiopoulos, A. G. Dimakis, R. Vadali, S. Chen, and D. Borthakur, "Xoring elephants: Novel erasure codes for big data," *Proceedings of the VLDB Endowment (to appear)*, 2013.
- [15] N. Silberstein, A. S. Rawat, O. O. Koiluoglu, and S. Vishwanath, "Optimal locally repairable codes via rank-metric codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Istanbul, 2013, pp. 1819–1823.
- [16] W. Song, S. Dau, C. Yuen, and T. Li, "Optimal locally repairable linear codes," *IEEE J. Sel. Areas Commun.*, vol. 32, pp. 6925–6934, May 2014.
- [17] I. Tamo and A. Barg, "Bounds on locally recoverable codes with multiple recovering sets," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Honolulu, 2014, pp. 691–695.
- [18] I. Tamo, D. S. Papailiopoulos, and A. G. Dimakis, "Optimal locally repairable codes and connections to matroid theory," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Istanbul, 2013, pp. 1814–1818.
- [19] Itzhak Tamo and Alexander Barg, "A family of optimal locally recoverable codes," *IEEE Trans. on Inform. Theory*, vol. 60, pp. 4661–4676, Aug. 2014.
- [20] A. Wang and Z. Zhang, "Repair locality with multiple erasure tolerance," *arXiv preprint arXiv:1306.4774*, 2013.
- [21] A. Wang and Z. Zhang, "Repair locality from a combinatorial perspective," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Honolulu, 2014, pp. 1972–1976.

## APPENDIX A

### PROOF OF PROPOSITION 13

**Lemma 19.** For  $1 \leq x \leq n_1$ ,

$$\Psi(x) \geq xr + \max\left\{\left\lceil \frac{x}{\lambda + 1} \right\rceil, \left\lceil \frac{x - \nu}{\lambda} \right\rceil\right\}.$$

*Proof:* Set

$$\begin{cases} s = \mu, \\ t_1 = \cdots = t_\nu = \lambda + 1, t_{\nu+1} = \cdots = t_\mu = \lambda, \\ a_1 = \cdots = a_\nu = \lambda, a_{\nu+1} = \cdots = a_\mu = \lambda - 1. \end{cases}$$

It is clear that  $s$  and  $\{t_i, a_i\}_{i \in [s]}$  satisfy (7), and then we have

$$\begin{aligned} \Psi(x) &\geq \underset{l, h_1, \dots, h_l}{\text{Min}} (xr + 1 - \sum_{i=1}^{l-1} (a_{h_i} - t_{h_i})) \\ &= \underset{l, h_1, \dots, h_l}{\text{Min}} (xr + l), \end{aligned}$$

where the minimum is subject to (8). On the other hand, for any integers  $l, h_1, \dots, h_l$  satisfying (8),

$$x \leq t_{h_1} + \cdots + t_{h_l} \leq \begin{cases} (\lambda + 1)l, & \text{if } l \leq \nu, \\ (\lambda + 1)\nu + (l - \nu)\lambda, & \text{if } l > \nu, \end{cases}$$

which induces  $x \leq \min\{\lambda l + l, \lambda l + \nu\}$ . Therefore  $l \geq \max\{\lceil \frac{x}{\lambda+1} \rceil, \lceil \frac{x-\nu}{\lambda} \rceil\}$  and then

$$\Psi(x) \geq \underset{l, h_1, \dots, h_l}{\text{Min}} (xr + l) \geq xr + \max\left\{\left\lceil \frac{x}{\lambda+1} \right\rceil, \left\lceil \frac{x-\nu}{\lambda} \right\rceil\right\}.$$

■

**Lemma 20.** For  $1 \leq x \leq n_1$ ,

$$\Psi(x) \leq xr + \max\left\{\left\lceil \frac{x}{\lambda+1} \right\rceil, \left\lceil \frac{x-\nu}{\lambda} \right\rceil\right\}.$$

*Proof:* We prove the lemma by contradiction. Assume that for some  $1 \leq x \leq n_1$ ,

$$\Psi(x) \geq xr + 1 + \max\left\{\left\lceil \frac{x}{\lambda+1} \right\rceil, \left\lceil \frac{x-\nu}{\lambda} \right\rceil\right\}.$$

Then there exist integers  $s$  and  $t_i, a_i$ ,  $1 \leq i \leq s$ , satisfying the constraints (7) and

$$\underset{l, h_1, \dots, h_l}{\text{Min}} (xr + 1 - \sum_{i=1}^{l-1} (a_{h_i} - t_{h_i})) \geq xr + 1 + \max\left\{\left\lceil \frac{x}{\lambda+1} \right\rceil, \left\lceil \frac{x-\nu}{\lambda} \right\rceil\right\},$$

where the minimum is subject to the constraint (8). Therefore for all integers  $l$  and  $h_1, \dots, h_l \in [s]$  satisfying the constraint (8), it has

$$\sum_{i=1}^{l-1} (a_{h_i} - t_{h_i}) \leq -\max\left\{\left\lceil \frac{x}{\lambda+1} \right\rceil, \left\lceil \frac{x-\nu}{\lambda} \right\rceil\right\}. \quad (15)$$

Consider the following two cases.

**Case 1.**  $1 \leq x \leq (\lambda + 1)\nu$ . Then  $\max\{\lceil \frac{x}{\lambda+1} \rceil, \lceil \frac{x-\nu}{\lambda} \rceil\} = \lceil \frac{x}{\lambda+1} \rceil$ . For  $1 \leq i \leq s$ , define

$$b_i = (\lambda + 1)a_i - \lambda t_i.$$

Then without loss of generality, we can assume that  $b_1 \geq b_2 \geq \cdots \geq b_s$ . Let  $h$  be the smallest integer such that  $1 \leq h \leq s$  and  $\sum_{i=1}^h (a_i - t_i) \leq -\lceil \frac{x}{\lambda+1} \rceil$ . Note that  $h$  exists because  $\sum_{i=1}^s (a_i - t_i) = n_2 - n_1 = -\mu \leq -\lceil \frac{x}{\lambda+1} \rceil$ . Next we consider the value of  $t_1 + \cdots + t_h$ .

If  $t_1 + \cdots + t_h \geq x$ , there exists a positive integer  $h' \leq h$  such that  $\sum_{j=1}^{h'-1} t_j < x \leq \sum_{j=1}^{h'} t_j$ . Then  $h' - 1 < h$  and by (15),

$$\sum_{i=1}^{h'-1} (a_i - t_i) \leq - \left\lceil \frac{x}{\lambda + 1} \right\rceil,$$

which contradicts to the minimality of  $h$ .

If  $t_1 + \cdots + t_h < x$ , we compute  $\sum_{i=1}^s b_i$  in two different ways. On the one hand,

$$\begin{aligned} \sum_{i=1}^s b_i &= \sum_{i=1}^s ((\lambda + 1)a_i - \lambda t_i) \\ &= (\lambda + 1)n_2 - \lambda n_1 \\ &= \nu - \mu. \end{aligned} \tag{16}$$

On the other hand, we claim that

- (i)  $\sum_{i=1}^h b_i \leq -1$  and  $b_i \leq -1$  for  $h + 1 \leq i \leq s$ ,
- (ii)  $h - s \leq \nu - \mu$ ,

and then

$$\begin{aligned} \sum_{i=1}^s b_i &= \left( \sum_{i=1}^h b_i \right) + \left( \sum_{i=h+1}^s b_i \right) \\ &\leq -1 + (-1) \times (s - h) \\ &= -1 + h - s \\ &\leq \nu - \mu - 1, \end{aligned}$$

which contradicts to (16).

In fact, the claim (i) holds because

$$\begin{aligned} \sum_{i=1}^h b_i &= \sum_{i=1}^h ((\lambda + 1)a_i - \lambda t_i) \\ &= (\lambda + 1) \sum_{i=1}^h (a_i - t_i) + \sum_{i=1}^h t_i \\ &\stackrel{(a)}{\leq} -(\lambda + 1) \left\lceil \frac{x}{\lambda + 1} \right\rceil + x - 1 \\ &\leq -1, \end{aligned}$$

where (a) follows from  $\sum_{i=1}^h (a_i - t_i) \leq -\left\lceil \frac{x}{\lambda + 1} \right\rceil$  and  $\sum_{i=1}^h t_i < x$ . Then  $b_j \leq \frac{1}{h} \sum_{i=1}^h b_i < 0$  for  $h + 1 \leq j \leq s$ . To show the claim (ii), observe that  $a_i \geq t_i - 1$  and  $\sum_{i=1}^{h-1} (a_i - t_i) \geq -\left\lceil \frac{x}{\lambda + 1} \right\rceil + 1$  from the minimality of  $h$ . Then we have

$$\begin{aligned} -\mu &= n_2 - n_1 = \sum_{i=1}^s (a_i - t_i) \\ &= \sum_{i=1}^{h-1} (a_i - t_i) + \sum_{i=h}^s (a_i - t_i) \\ &\geq -\left\lceil \frac{x}{\lambda + 1} \right\rceil + 1 + (-1) \times (s - h + 1). \end{aligned}$$

Because  $x \leq (\lambda + 1)\nu$ , it holds

$$-\mu \geq -\nu + 1 + (-1) \times (s - h + 1) = -\nu + h - s,$$

and the claim (ii) follows directly.

**Case 2.**  $(\lambda + 1)\nu + 1 \leq x \leq n_1$ . Then  $\max\{\lceil \frac{x}{\lambda+1} \rceil, \lceil \frac{x-\nu}{\lambda} \rceil\} = \lceil \frac{x-\nu}{\lambda} \rceil$ . Similar to Case 1, define

$$c_i = \lambda a_i - (\lambda - 1)t_i, \quad \forall 1 \leq i \leq s$$

and assume  $c_1 \geq c_2 \geq \dots \geq c_s$ . Let  $g \in [s]$  be the smallest positive integer such that  $\sum_{i=1}^g (a_i - t_i) \leq -\lceil \frac{x-\nu}{\lambda} \rceil$ . Note that  $g$  exists because  $\sum_{i=1}^s (a_i - t_i) = n_2 - n_1 = -\mu \leq -\lceil \frac{x-\nu}{\lambda} \rceil$ . Next we consider the value of  $t_1 + \dots + t_g$ .

Similar to Case 1,  $t_1 + \dots + t_g \geq x$  contradicts to the minimality of  $g$ . Then it follows  $t_1 + \dots + t_g < x$ . We compute the value of  $\sum_{i=1}^s c_i$  in two different ways. On the one hand,

$$\begin{aligned} \sum_{i=1}^s c_i &= \lambda \sum_{i=1}^s a_i - (\lambda - 1) \sum_{i=1}^s t_i \\ &= \lambda n_2 - (\lambda - 1)n_1 \\ &= \nu. \end{aligned} \tag{17}$$

On the other hand, we claim that

- (i)  $\sum_{i=1}^g c_i \leq \nu - 1$ ,
- (ii)  $c_i \leq 0$  for  $g + 1 \leq i \leq s$ .

Then

$$\begin{aligned} \sum_{i=1}^s c_i &= \sum_{i=1}^g c_i + \sum_{i=g+1}^s c_i \\ &\leq \sum_{i=1}^g c_i \leq \nu - 1, \end{aligned}$$

which contradicts to (17).

Note that  $\sum_{i=1}^g (a_i - t_i) \leq -\lceil \frac{x-\nu}{\lambda} \rceil$  and  $\sum_{i=1}^g t_i < x$ , then the claim (i) follows from

$$\begin{aligned} \sum_{i=1}^g c_i &= \lambda \sum_{i=1}^g (a_i - t_i) + \sum_{i=1}^g t_i \\ &\leq -\lambda \left\lceil \frac{x-\nu}{\lambda} \right\rceil + x - 1 \\ &\leq \nu - 1. \end{aligned}$$

To show the claim (ii), observe that  $c_j \leq \frac{1}{g} \sum_{i=1}^g c_i \leq \frac{\nu-1}{g}$  for  $g + 1 \leq j \leq s$ , and  $g \geq -\sum_{i=1}^g (a_i - t_i) \geq \lceil \frac{x-\nu}{\lambda} \rceil > \nu$  where the first inequality is from  $a_i \geq t_i - 1$  for  $1 \leq i \leq s$  and the last inequality is from  $x \geq (\lambda + 1)\nu + 1$ . Then it has  $c_j < \frac{\nu}{g} < 1$  for  $g + 1 \leq j \leq s$  and the claim (ii) then follows. ■

## APPENDIX B PROOF OF LEMMA 17

(i) Because  $X = (\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_r)$  generates an  $[r + 1, r]$  MDS code, there exist nonzero elements  $e_0, e_1, \dots, e_r \in \mathbb{F}_q$  such that  $e_0 \mathbf{x}_0 + e_1 \mathbf{x}_1 + \dots + e_r \mathbf{x}_r = 0$ . Moreover, since  $\mathbf{c} = (1, c_1, \dots, c_r)$  is a codeword of the MDS code, it has  $e_0 + e_1 c_1 + \dots + e_r c_r = 0$ . Therefore  $e_0 \alpha_0 + e_1 \alpha_{i,1} + \dots + e_r \alpha_{i,r} = 0$  for  $1 \leq i \leq \lambda + 1$ . Thus (i) follows directly.

(ii) We prove the statement by contradiction. Assume that the vectors in  $F$  are linearly dependent, i.e. there exists  $e_\alpha \in \mathbb{F}_q$  for each  $\alpha \in F$  such that  $\sum_{\alpha \in F} e_\alpha \alpha = 0$ , where  $\{e_\alpha\}_{\alpha \in F}$  are not all zeros. In fact,

at least two out of  $\{e_\alpha\}_{\alpha \in F}$  are nonzero because the vectors in  $F$  are not zero vectors. We consider the following two cases.

**Case 1.**  $|(F - \{\alpha_0\}) \cap \mathcal{A}_i| \leq r - 1$  for  $1 \leq i \leq \lambda + 1$ . Because at least two out of  $\{e_\alpha\}_{\alpha \in F}$  are nonzero, there exists  $i_0 \in [\lambda + 1]$  such that the coefficients  $\{e_\alpha\}_{\alpha \in \mathcal{A}_{i_0} \setminus \{\alpha_0\}}$  are not all zero. Then without loss of generality, assume  $(F - \{\alpha_0\}) \cap \mathcal{A}_{i_0} = \{\alpha_{i_0,1}, \alpha_{i_0,1} \dots, \alpha_{i_0,h}\}$ , where  $h \leq r - 1$ . Consider the restriction of the linear combination  $\sum_{\alpha \in F} e_\alpha \alpha$  to its  $i_0$ th thick row, (i.e., the  $((i_0 - 1)r + 1)$ -th row to the  $i_0 r$ -th row,) we have  $\sum_{j=1}^h e_{\alpha_{i_0,j}} \mathbf{x}_j = e \mathbf{x}_0$  for some  $e \in \mathbb{F}_q$ . It follows that  $\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_h$  are  $\mathbb{F}_q$ -linearly dependent, where  $h \leq r - 1$ , which contradicts the fact that  $(\mathbf{x}_0, \dots, \mathbf{x}_r)$  generates an  $[r + 1, r]$  MDS code.

**Case 2.** For some  $(i_0, j_0) \in [\lambda + 1] \times [r]$ ,  $|(F - \{\alpha_{i_0,j_0}\}) \cap \mathcal{A}_i| \leq r - 1$  for  $1 \leq i \leq \lambda + 1$ . Without loss of generality, assume  $i_0 = j_0 = 1$ , i.e.,  $|(F - \{\alpha_{1,1}\}) \cap \mathcal{A}_i| \leq r - 1$  for  $1 \leq i \leq \lambda + 1$ . If there exists  $l$ ,  $2 \leq l \leq \lambda + 1$ , such that  $\{e_\alpha\}_{\alpha \in \mathcal{A}_l \setminus \{\alpha_0\}}$  are not all zero, then similar to Case 1, restricting the linear combination  $\sum_{\alpha \in F} e_\alpha \alpha$  to its  $l$ th thick row will lead a contradiction. Therefore we have  $e_\alpha = 0$  for all  $\alpha \in \cup_{i=2}^{\lambda+1} \mathcal{A}_i - \{\alpha_0\}$ . Thus it suffice to check the vectors in  $F \cap \mathcal{A}_1$ . Similarly, a contradiction arises when restricting  $\sum_{\alpha \in F} e_\alpha \alpha$  to the first thick row.

## APPENDIX C PROOF OF THE CLAIM

**Lemma 21.** *For any  $V \subseteq \Omega$  with  $|V| = k + \tilde{\eta}$ , there exist subsets  $V_1, \dots, V_\mu \subseteq V$  such that the following two conditions are satisfied:*

- (1)  $|\cup_{l=1}^\mu V_l| \geq k$ ;
- (2) For  $1 \leq l \leq \mu$ ,  $V_l \subseteq \cup_{i=1}^{\lambda+1} W_i^{(l)}$ , and there exists  $\omega_l \in V_l$  such that  $|(V_l - \{\omega_l\}) \cap W_i^{(l)}| \leq r - 1$  for all  $i \in [\lambda + 1]$ .

*Proof:* Denote  $U_l = V \cap (\cup_{i=1}^{\lambda+1} W_i^{(l)})$  for  $1 \leq l \leq \mu$ . Then the proof is completed by two steps. First, we show that for all nonempty set  $U_l$ ,  $1 \leq l \leq \mu$ , there exists a subset  $V_l \subseteq U_l$  satisfying

- $|V_l| \geq |U_l| - \left\lfloor \frac{|U_l|-1}{r} \right\rfloor$ ; and
- There exists  $\omega_l \in V_l$  such that  $|(V_l - \{\omega_l\}) \cap W_i^{(l)}| \leq r - 1$  for all  $i \in [\lambda + 1]$ .

Second, by setting  $V_l = \emptyset$  for all  $l \in [\mu]$  with  $|U_l| = 0$ , we prove that  $|V_1 \cup V_2 \cup \dots \cup V_\mu| \geq k$ . The details are given below.

**Step 1.** Suppose  $U_l$  is nonempty. Consider the following two cases.

(a)  $\omega_0^{(l)} \in U_l$ . Then there are at most  $\left\lfloor \frac{|U_l|-1}{r} \right\rfloor$  sets out of  $W_1^{(l)}, W_2^{(l)}, \dots, W_{\lambda+1}^{(l)}$  which are contained in  $U_l$ , say,  $W_1^{(l)}, \dots, W_h^{(l)} \subseteq U_l$ , where  $h \leq \left\lfloor \frac{|U_l|-1}{r} \right\rfloor$ . Define  $V_l$  by deleting  $\omega_{1,1}^{(l)}, \omega_{2,1}^{(l)}, \dots, \omega_{h,1}^{(l)}$  from  $U_l$ , then we have  $|(V_l - \{\omega_0^{(l)}\}) \cap W_i^{(l)}| \leq r - 1$  for all  $i \in [\lambda + 1]$  and  $|V_l| \geq |U_l| - \left\lfloor \frac{|U_l|-1}{r} \right\rfloor$ .

(b)  $\omega_0^{(l)} \notin U_l$ . Similarly, there are at most  $\left\lfloor \frac{|U_l|}{r} \right\rfloor$  sets out of  $W_1^{(l)}, W_2^{(l)}, \dots, W_{\lambda+1}^{(l)}$  which are contained in  $U_l \cup \{\omega_0^{(l)}\}$ , say,  $W_1^{(l)}, \dots, W_{h'}^{(l)} \subseteq U_l \cup \{\omega_0^{(l)}\}$ , where  $h' \leq \left\lfloor \frac{|U_l|}{r} \right\rfloor$ . Define  $V_l$  by deleting  $\omega_{2,1}^{(l)}, \omega_{3,1}^{(l)}, \dots, \omega_{h',1}^{(l)}$  from  $U_l$ , then we have  $|(V_l - \{\omega_{1,1}^{(l)}\}) \cap W_i^{(l)}| \leq r - 1$  for all  $i \in [\lambda + 1]$ , and  $|V_l| \geq |U_l| - \left( \left\lfloor \frac{|U_l|}{r} \right\rfloor - 1 \right) \geq |U_l| - \left\lfloor \frac{|U_l|-1}{r} \right\rfloor$ .

**Step 2.** Observe that

$$\begin{aligned}
|\cup_{l=1}^{\mu} V_l| &= \sum_{l \in [\mu], U_l \neq \emptyset} |V_l| \\
&\geq \sum_{l \in [\mu], U_l \neq \emptyset} (|U_l| - \left\lfloor \frac{|U_l| - 1}{r} \right\rfloor) \\
&= k + \tilde{\eta} - \sum_{l \in [\mu], U_l \neq \emptyset} \left\lfloor \frac{|U_l| - 1}{r} \right\rfloor \\
&\geq k + \tilde{\eta} - \left\lfloor \sum_{l \in [\mu], U_l \neq \emptyset} \frac{|U_l| - 1}{r} \right\rfloor \\
&= k + \tilde{\eta} - \left\lfloor \frac{k + \tilde{\eta} - \epsilon}{r} \right\rfloor,
\end{aligned}$$

where  $\epsilon = |\{l \in [l] : U_l \neq \emptyset\}|$ . Then it suffices to show  $\left\lfloor \frac{k + \tilde{\eta} - \epsilon}{r} \right\rfloor \leq \tilde{\eta}$ .

Denote  $\epsilon_1 = |\{l : 1 \leq l \leq \nu, U_l \neq \emptyset\}|$  and  $\epsilon_2 = |\{l : \nu + 1 \leq l \leq \mu, U_l \neq \emptyset\}|$ , then  $\epsilon = \epsilon_1 + \epsilon_2$ . Because  $|U_1| + |U_2| + \dots + |U_\mu| = k + \tilde{\eta}$  and

$$|U_l| \leq \begin{cases} |\cup_{i=1}^{\lambda+1} W_i^{(l)}| = (\lambda + 1)r + 1, & \text{for } 1 \leq l \leq \nu \\ |\cup_{i=1}^{\lambda} W_i^{(l)}| = \lambda r + 1, & \text{for } \nu + 1 \leq l \leq \mu, \end{cases}$$

we have

$$\begin{cases} 0 \leq \epsilon_1 \leq \nu; \\ 0 \leq \epsilon_2 \leq \mu - \nu; \\ k + \tilde{\eta} \leq \epsilon_1((\lambda + 1)r + 1) + \epsilon_2(\lambda r + 1). \end{cases}$$

Since  $\epsilon_1((\lambda + 1)r + 1) + \epsilon_2(\lambda r + 1) \leq ((\lambda + 1)r + 1)(\epsilon_1 + \epsilon_2)$  and also  $\epsilon_1((\lambda + 1)r + 1) + \epsilon_2(\lambda r + 1) = (\lambda r + 1)(\epsilon_1 + \epsilon_2) + \epsilon_1 r \leq (\lambda r + 1)(\epsilon_1 + \epsilon_2) + \nu r$ , it follows that  $\epsilon \geq \max\{\frac{k + \tilde{\eta}}{(\lambda + 1)r + 1}, \frac{k + \tilde{\eta} - \nu r}{\lambda r + 1}\}$ . Thus

$$\begin{aligned}
\left\lfloor \frac{k + \tilde{\eta} - \epsilon}{r} \right\rfloor &\leq \left\lfloor \frac{1}{r} (k + \tilde{\eta} - \max\{\frac{k + \tilde{\eta}}{(\lambda + 1)r + 1}, \frac{k + \tilde{\eta} - \nu r}{\lambda r + 1}\}) \right\rfloor \\
&= \left\lfloor \frac{1}{r} \min\{\frac{(k + \tilde{\eta})(\lambda + 1)r}{(\lambda + 1)r + 1}, \frac{(k + \tilde{\eta})\lambda r + \nu r}{\lambda r + 1}\} \right\rfloor \\
&= \min\left\{ \left\lfloor \frac{(k + \tilde{\eta})(\lambda + 1)}{(\lambda + 1)r + 1} \right\rfloor, \left\lfloor \frac{(k + \tilde{\eta})\lambda + \nu}{\lambda r + 1} \right\rfloor \right\}.
\end{aligned}$$

Note that  $\tilde{\eta} = \min\left\{ \left\lceil \frac{(\lambda + 1)(k - 1) + 1}{(\lambda + 1)(r - 1) + 1} \right\rceil, \left\lceil \frac{\lambda(k - 1) + \nu + 1}{\lambda(r - 1) + 1} \right\rceil \right\} - 1$ . Then if  $\tilde{\eta} = \left\lceil \frac{(\lambda + 1)(k - 1) + 1}{(\lambda + 1)(r - 1) + 1} \right\rceil - 1$ , it has

$$\begin{aligned}
\frac{(k + \tilde{\eta})(\lambda + 1)}{(\lambda + 1)r + 1} - (\tilde{\eta} + 1) &= \frac{(\lambda + 1)(k - 1) - ((\lambda + 1)(r - 1) + 1)(\tilde{\eta} + 1)}{(\lambda + 1)r + 1} \\
&= \frac{(\lambda + 1)(r - 1) + 1}{(\lambda + 1)r + 1} \times \left( \frac{(\lambda + 1)(k - 1)}{(\lambda + 1)(r - 1) + 1} - (\tilde{\eta} + 1) \right) \\
&< 0,
\end{aligned}$$

and therefore  $\left\lfloor \frac{(k + \tilde{\eta})(\lambda + 1)}{(\lambda + 1)r + 1} \right\rfloor \leq \tilde{\eta}$ . Similarly, if  $\tilde{\eta} = \left\lceil \frac{\lambda(k - 1) + \nu + 1}{\lambda(r - 1) + 1} \right\rceil - 1$ , it can be proved that  $\left\lfloor \frac{(k + \tilde{\eta})\lambda + \nu}{\lambda r + 1} \right\rfloor \leq \tilde{\eta}$ . Thus we conclude that  $\left\lfloor \frac{k + \tilde{\eta} - \epsilon}{r} \right\rfloor \leq \tilde{\eta}$ . ■